

INFERENCE IN A BIMODAL BIRNBAUM-SAUNDERS MODEL

RODNEY FONSECA AND FRANCISCO CRIBARI-NETO

ABSTRACT. We address the issue of performing inference on the parameters that index a bimodal extension of the Birnbaum-Saunders distribution (\mathcal{BS}). We show that maximum likelihood point estimation can be problematic since the standard nonlinear optimization algorithms may fail to converge. To deal with this problem, we penalize the log-likelihood function. The numerical evidence we present show that maximum likelihood estimation based on such penalized function is made considerably more reliable. We also consider hypothesis testing inference based on the penalized log-likelihood function. In particular, we consider likelihood ratio, signed likelihood ratio, score and Wald tests. Bootstrap-based testing inference is also considered. We use a nonnested hypothesis test to distinguish between two bimodal \mathcal{BS} laws. We derive analytical corrections to some tests. Monte Carlo simulation results and empirical applications are presented and discussed.

KEYWORDS: Bimodal Birnbaum-Saunders distribution; Birnbaum-Saunders distribution; monotone likelihood; nonnested hypothesis test; penalized likelihood; signed likelihood ratio.

1. INTRODUCTION

The Birnbaum-Saunders distribution was proposed by Birnbaum and Saunders (1969a) to model failure time due to fatigue under cyclic loading. In such a model, failure follows from the development and growth of a dominant crack. Based on that setup, the authors obtained the following distribution function:

$$F(x) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}} \right) \right], \quad x > 0, \quad (1)$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively, and $\Phi(\cdot)$ is the standard normal distribution function. We write $X \sim \mathcal{BS}(\alpha, \beta)$.

Maximum likelihood estimation of the parameters that index the \mathcal{BS} distribution was first investigated by Birnbaum and Saunders (1969b). Bias-corrected estimators were obtained by Lemonte et al. (2007) and Lemonte et al. (2008). Improved maximum likelihood estimation of the \mathcal{BS} parameters was developed by Cysneiros et al. (2008). Ng et al. (2003) compared the finite-sample performance of maximum likelihood estimators (MLEs) to that of estimators obtained using the modified method of moments. For details on the \mathcal{BS} distribution, its main properties and applications, readers are referred to Leiva (2015).

Several extensions of the \mathcal{BS} distribution have been proposed in the literature aiming at making the model more flexible. For instance, Díaz-García and Leiva (2005) and Sanhueza et al. (2008) used non-Gaussian kernels to extend the \mathcal{BS} model. The \mathcal{BS} distribution was also extended through the inclusion of additional parameters; see, e.g., Díaz-García and Dominguez-Molina (2006), Owen (2006) and Owen and Ng (2015). More recently, extensions of the \mathcal{BS} model were proposed by Bourguignon et al. (2014), Cordeiro et al. (2013), Cordeiro and Lemonte (2014) and Zhu and Balakrishnan

(2015). Alternative approaches are the use of scale-mixture of normals, as discussed by Balakrishnan et al. (2009) and Patriota (2012), for example, and the use of mixtures of \mathcal{BS} distributions, as in Balakrishnan et al. (2011). Again, details can be found in Leiva (2015).

A bimodal \mathcal{BS} distribution was proposed by Olmos et al. (2016). The authors used the approach described in Gómez et al. (2011) to obtain a variation of the \mathcal{BS} model that can assume bimodal shapes. Another variant of the \mathcal{BS} distribution that exhibits bimodality was discussed by Díaz-García and Dominguez-Molina (2006) and Owen and Ng (2015), which the latter authors denoted by \mathcal{GBS}_2 . In their model, bimodality takes place when two parameter values exceed certain thresholds. In what follows we shall work with the \mathcal{BBS} model instead of the \mathcal{GBS}_2 distribution because in the former bimodality is controlled by a single parameter. Even though we shall focus on the \mathcal{BBS} distribution, in some parts of the paper we shall consider the \mathcal{GBS}_2 law as an alternative model; see Section 6 for further details.

A problem with the \mathcal{BBS} distribution we detected is that log-likelihood maximizations based on Newton or quasi-Newton methods oftentimes fail to converge. In this paper we analyze some possible solutions to such a problem, such as the use of resampling methods and the inclusion of a penalization term in the log-likelihood function.

As a motivation, consider the data provided by Folks and Chhikara (1978) that consist of 25 observations on runoff amounts at Jug Bridge, in Maryland. Figure 1a shows log-likelihood contour curves obtained by varying the values of α and γ while keeping the value of β fixed. Notice that there is a region apparently flat of the profile log-likelihood function, which might be making the optimization process to fail to converge. In Figure 1b we present similar contour curves for a penalized version of the log-likelihood function. It can be seen that plausible estimates are obtained. We shall return to this application in Section 7.

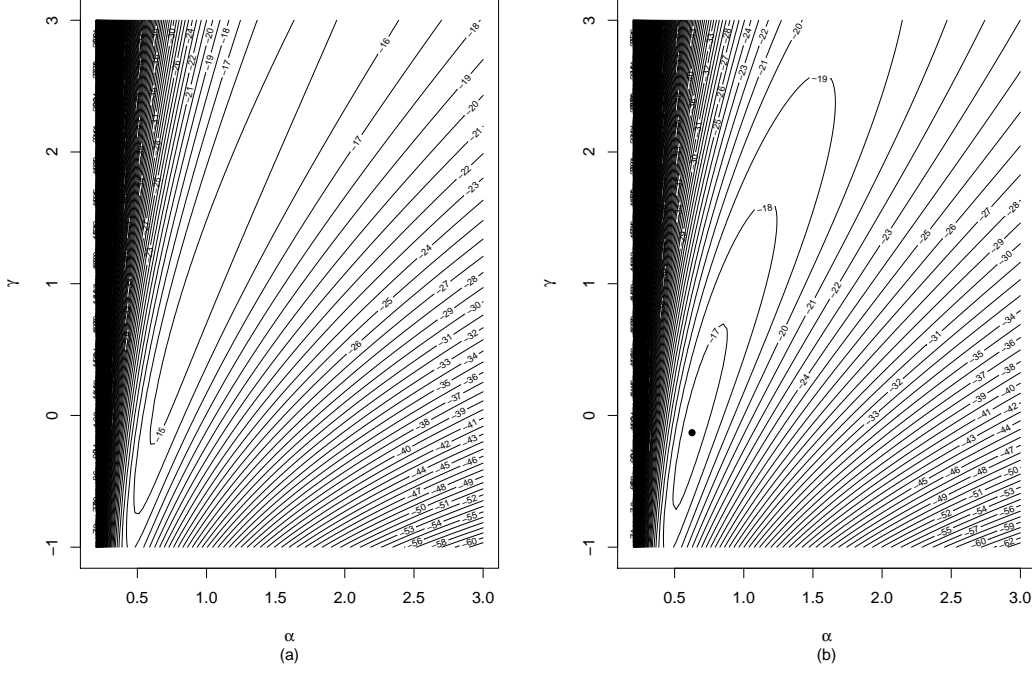
The chief goal of our paper is to provide a solution to the convergence failure and implausible parameter estimates associated with log-likelihood maximization in the \mathcal{BBS} model. We compare different estimation procedures and propose to include a penalization term in the log-likelihood function. In particular, regions of the parameter space where the likelihood is flat or nearly flat are heavily penalized. That approach considerably improves maximum likelihood parameter estimation. We also focus on hypothesis testing inference based on the penalized log-likelihood function. For instance, a one-sided hypothesis test is used to test whether the variate follows the \mathcal{BBS} law with two modes. Analytical and bootstrap corrections are proposed to improve the finite sample performances such test. Moreover, we present nonnested hypothesis tests that can be used to distinguish between two bimodal extensions of the \mathcal{BS} distribution, the \mathcal{BBS} and \mathcal{GBS}_2 models. The finite sample performances of all tests are numerically evaluated using Monte Carlo simulations.

The paper unfolds as follows. Section 2 presents the \mathcal{BBS} distribution and its main properties. Simulation results are presented in Section 3, where we outline some possible solutions to the numerical difficulties associated with \mathcal{BBS} log-likelihood maximization. Two-sided hypothesis tests in the \mathcal{BBS} model are discussed in Section 4. In Section 5 we focus on one-sided tests where the main interest lies in detecting bimodality. Section 6 describes nonnested hypothesis testing inference. Empirical applications are presented and discussed in Section 7. Finally, some concluding remarks are offered in Section 8.

2. THE BIMODAL BIRNBAUM-SAUNDERS DISTRIBUTION

The Birnbaum-Saunders distribution proposed by Olmos et al. (2016) can be used to model positive data and is more flexible than the original \mathcal{BS} distribution since it can accommodate bimodality. A

FIGURE 1. Contour curves of the profile log-likelihood of α and γ with $\beta = 0.69$ (fixed) for the runoff amounts data. Panel (a) corresponds to no penalization and panel (b) follows from penalizing the log-likelihood function.



random variable X is $\mathcal{BS}(\alpha, \beta, \gamma)$ distributed if its probability density function (pdf) is given by

$$f(x) = \frac{x^{-3/2}(x + \beta)}{4\alpha\beta^{1/2}\Phi(-\gamma)}\phi(|t| + \gamma), \quad x > 0, \quad (2)$$

where $\alpha, \beta > 0, \gamma \in \mathbb{R}$, $t = \alpha^{-1}(\sqrt{x/\beta} - \sqrt{\beta/x})$ and $\phi(\cdot)$ is the standard normal pdf. Figure 2 shows plots of the density in (2) for some parameter values. We note that when $\gamma < 0$ the density is bimodal.

The cumulative distribution function (cdf) of X is

$$F(x) = \left[\frac{\Phi(t - \gamma)}{2\Phi(-\gamma)} \right]^{I(x, \beta)} \left[\frac{1}{2} + \frac{\Phi(t - \gamma) - \Phi(\gamma)}{2\Phi(-\gamma)} \right]^{1 - I(x, \beta)}, \quad x > 0, \quad (3)$$

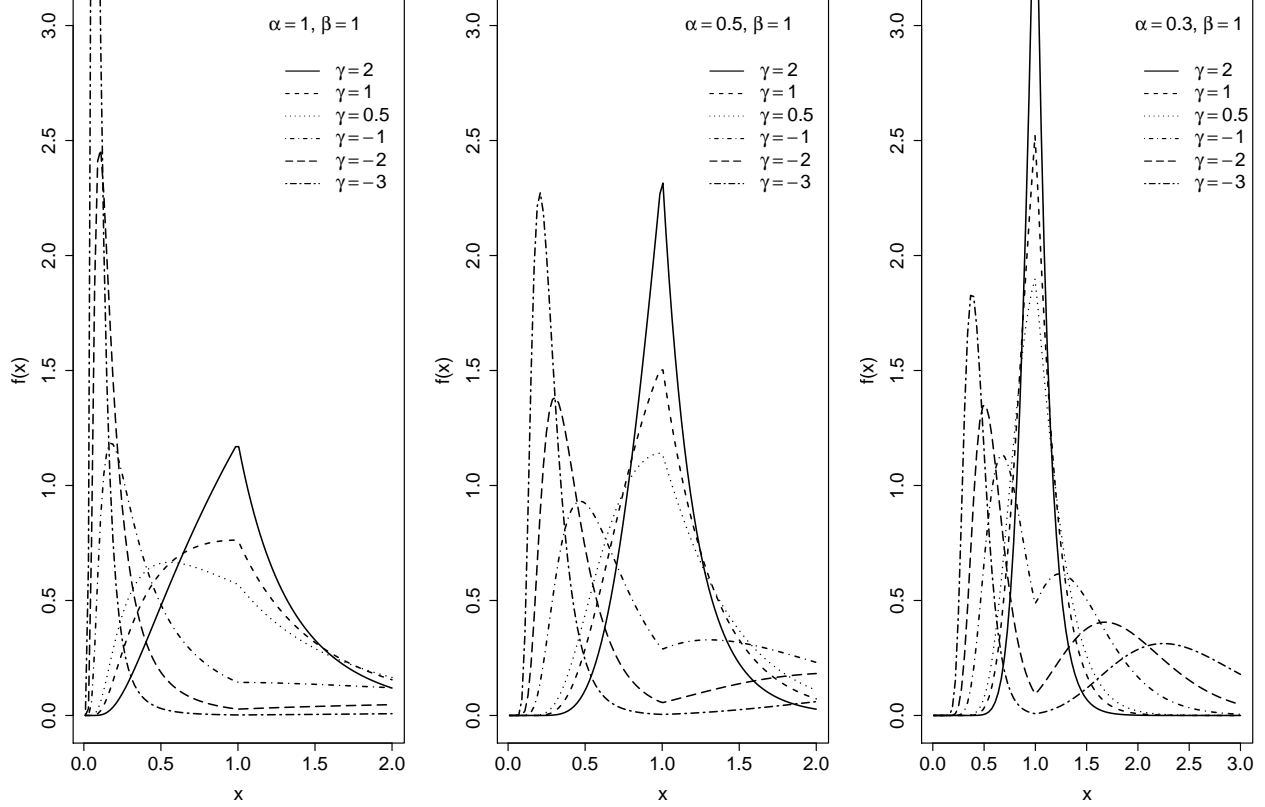
where

$$I(x, \beta) = \begin{cases} 1 & \text{if } x < \beta \\ 0 & \text{if } x \geq \beta \end{cases},$$

and $\Phi(\cdot)$ denotes the standard normal cdf. Some key properties of the \mathcal{BS} distribution also hold for \mathcal{BS} model, such as proportionality and reciprocity closure, i.e., $aX \sim \mathcal{BS}(\alpha, a\beta, \gamma)$ and $X^{-1} \sim \mathcal{BS}(\alpha, \beta^{-1}, \gamma)$, respectively, where a is a positive scalar.

An expression for the r th ordinary moment of X is

$$\mathbb{E}(X^r) = \frac{\beta^r}{\Phi(-\gamma)} \sum_{k=0}^r \sum_{j=0}^k \sum_{s=0}^m \binom{2r}{2k} \binom{k}{j} \binom{m}{s} \left(\frac{\alpha}{2}\right)^m (-\gamma)^{m-s} d_s(\gamma), \quad (4)$$

FIGURE 2. $\mathcal{BB}\mathcal{S}(\alpha, \beta, \gamma)$ densities for some parameter values.

where $r \in \mathbb{N}$ and $d_a(r)$ is the r th standard normal incomplete moment:

$$d_r(a) = \int_a^\infty t^r \phi(t) dt.$$

A useful stochastic representation is $Y = |T| + \gamma$. Here, Y follows the truncated standard normal distribution with support in (γ, ∞) , $T = (\sqrt{X/\beta} - \sqrt{\beta/X})/\alpha$ and $X \sim \mathcal{BB}\mathcal{S}(\alpha, \beta, \gamma)$. This relationship can be used to compute moments of the $\mathcal{BB}\mathcal{S}$ distribution.

3. LOG-LIKELIHOOD FUNCTIONS

Consider a row vector $\mathbf{x} = (x_1, \dots, x_n)$ of independent and identically distributed (iid) observations from the $\mathcal{BB}\mathcal{S}(\alpha, \beta, \gamma)$ distribution. Let $\theta = (\alpha, \beta, \gamma)$ be the vector of unknown parameters to be estimated. The log-likelihood function is

$$\begin{aligned} \ell(\theta) = & -n \log \{4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2}\} - \frac{3}{2} \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(x_i + \beta) \\ & - \frac{1}{2} \sum_{i=1}^n (|t_i| + \gamma)^2. \end{aligned} \tag{5}$$

Differentiating the log-likelihood function with respect to each parameter we obtain the score function $U_\theta = (U_\alpha, U_\beta, U_\gamma)$, where

$$U_\alpha = \frac{\partial \ell(\theta)}{\partial \alpha} = -\frac{n}{\alpha} + \frac{1}{2} \sum_{i=1}^n t_i^2 + \frac{\gamma}{\alpha} \sum_{i=1}^n |t_i|, \quad (6)$$

$$U_\beta = \frac{\partial \ell(\theta)}{\partial \beta} = -\frac{n}{2\beta} + \sum_{i=1}^n \frac{1}{x_i + \beta} + \sum_{i=1}^n \frac{\text{sign}(t_i)(|t_i| + \gamma)}{2\alpha\beta^{3/2}} \left(x_i^{1/2} + \frac{\beta}{x_i^{1/2}} \right), \quad (7)$$

$$U_\gamma = \frac{\partial \ell(\theta)}{\partial \gamma} = n \frac{\phi(\gamma)}{\Phi(-\gamma)} - n\gamma - \sum_{i=1}^n |t_i|, \quad (8)$$

and $\text{sign}(\cdot)$ represents the sign function.

The parameters maximum likelihood estimators, namely $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, can be obtained by solving $U_\theta = 0$. They cannot be expressed in closed-form and parameter estimates are obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton algorithm. To that end, one must specify an initial point for the iterative scheme. We propose using as starting values for α and β their modified method of moments estimates (Ng et al., 2003), and also using $\gamma = 0$ as a starting value; the latter means that the algorithm starts at the \mathcal{BS} law. We used such starting values in the numerical evaluations, and they proved to work well.

Based on several numerical experiments we noted a serious shortcoming: iterative numerical maximization of the \mathcal{BS} log-likelihood function may fail to converge and may yield implausible parameter estimates. Indeed, that is very likely to happen, especially when $\gamma > 0$. It is not uncommon to obtain very large (thus implausible) \mathcal{BS} parameter estimates, which is indicative that the likelihood function may be monotone; see Pianto and Cribari-Neto (2011). We shall address this problem in the subsections that follow.

3.1. Log-likelihood function penalized by the Jeffreys prior. An interesting estimation procedure was proposed by Firth (1993), where the score function is modified in order to reduce the bias of the maximum likelihood estimator. An advantage of this method is that maximum likelihood estimates need not be finite since the correction is applied in a preventive fashion. For models in the canonical exponential family, the correction can be applied directly to likelihood function:

$$L^*(\theta|\mathbf{x}) = L(\theta|\mathbf{x})|K|^{1/2},$$

where $|K|$ is the determinant of the expected information matrix. Thus, penalization of the likelihood function entails multiplying the likelihood function by the Jeffreys invariant prior.

Even though the \mathcal{BS} distribution is not a member of the canonical exponential family, we shall consider the above penalization scheme. In doing so, we follow Pianto and Cribari-Neto (2011) who used the same approach in speckled imagery analysis. We seek to prevent cases of monotone likelihood function that might lead to frequent optimization nonconvergences and implausible estimates. The $\mathcal{BS}(\alpha, \beta, \gamma)$ expected information matrix was obtained by Olmos et al. (2016). Its determinant is

$$|K| = \left[L_{\beta\beta} + \frac{1}{\alpha^2\beta^2} + \frac{\gamma(\gamma - \omega)}{4\beta^2} \right] \left[\frac{(\gamma - \omega)\omega(3 - \gamma\omega - \gamma^2) + 2}{\alpha^2} \right],$$

where $\omega = \phi(\gamma)/\Phi(-\gamma)$ and $L_{\beta\beta} = \mathbb{E}[(X + \beta)^{-2}]$. Thus, the log-likelihood function penalized by the Jeffreys prior can be written as

$$\ell^*(\theta) = -n \log \{ 4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2} \} - \frac{3}{2} \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(x_i + \beta)$$

$$\begin{aligned}
& -\frac{1}{2} \sum_{i=1}^n (t_i^2 + 2|t_i|\gamma + \gamma^2) + \frac{1}{2} \log \left[L_{\beta\beta} + \frac{1}{\alpha^2 \beta^2} + \frac{\gamma(\gamma - \omega)}{4\beta^2} \right] \\
& + \frac{1}{2} \log \left[\frac{(\gamma - \omega)\omega(3 + \gamma(\gamma - \omega)) + 2}{\alpha^2} \right].
\end{aligned} \tag{9}$$

If the likelihood function is monotone, the function becomes very flat for large parameter values and the Jeffreys penalization described above essentially eliminates such parameter range from the estimation. The likelihood of nonconvergences taking place and implausible estimates being obtained should be greatly reduced.

3.2. Log-likelihood function modified by the better bootstrap. An alternative approach uses the method proposed by Cribari-Neto et al. (2002), where bootstrap samples are used to improve maximum likelihood estimation similarly to the approach introduced by Efron (1990) and known as ‘the better bootstrap’. The former, however, does not require the estimators to have closed-form expressions. Based on the sample $\mathbf{x} = (x_1, \dots, x_n)$ of n observations, we obtain pseudo-samples \mathbf{x}^* of the same size by sampling from \mathbf{x} with replacement. Let P_i^* denote the proportion of times that observation x_i is selected, $i = 1, \dots, n$. We obtain the row vector $\mathbf{P}^{*b} = (P_1^{*b}, \dots, P_n^{*b})$ for the b th pseudo-sample, $b = 1, \dots, B$. Now compute

$$\mathbf{P}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \mathbf{P}_n^{*b},$$

i.e., compute the vector of mean selection frequencies using the B bootstrap samples. The vector $\mathbf{P}^*(\cdot)$ is then used to modify the log-likelihood function in the following manner:

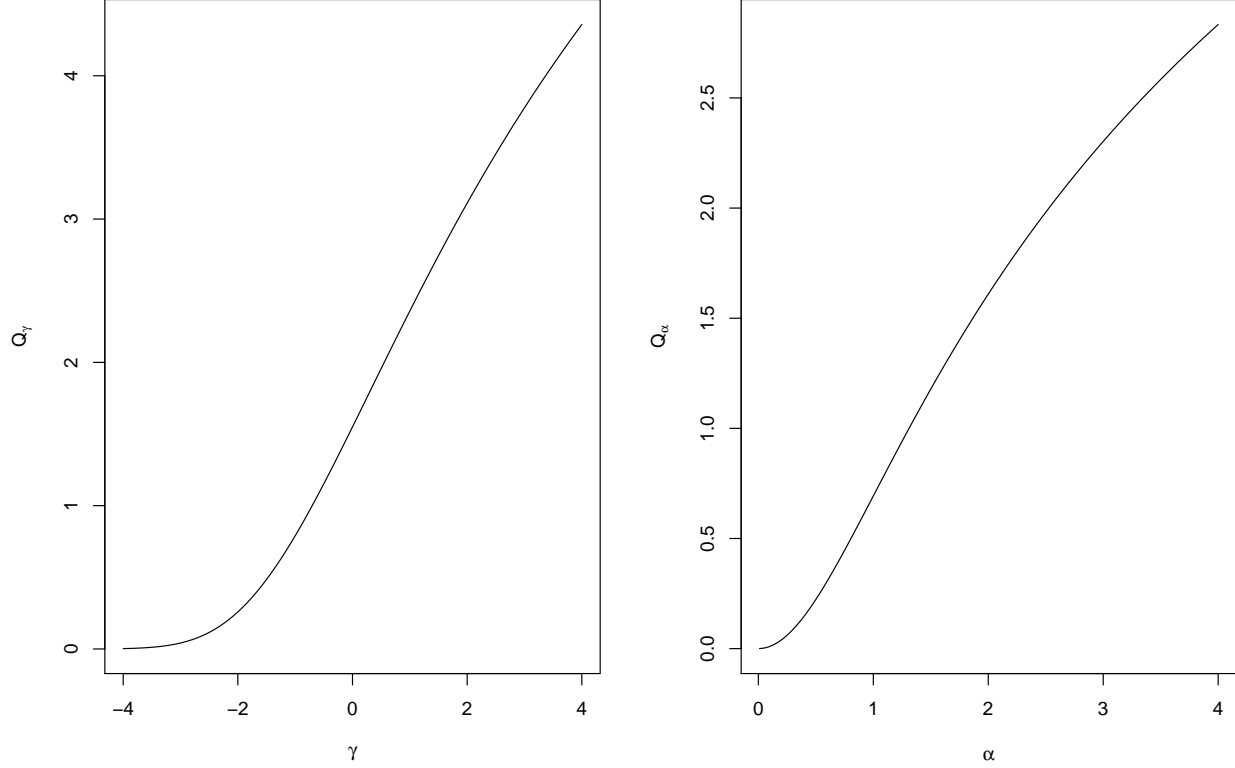
$$\ell(\theta) = -n \log \{4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2}\} - \frac{3n}{2} \mathbf{P}^*(\cdot) \log(\mathbf{x})^\top + n \mathbf{P}^*(\cdot) \log(\mathbf{x} + \beta)^\top - \frac{n}{2} \mathbf{P}^*(\cdot) \mathbf{t}_\gamma^\top,$$

where $\log(\mathbf{x}) = (\log(x_1), \dots, \log(x_n))$, $\log(\mathbf{x} + \beta) = (\log(x_1 + \beta), \dots, \log(x_n + \beta))$ and $\mathbf{t}_\gamma = ((|t_1| + \gamma)^2, \dots, (|t_n| + \gamma)^2)$ are row vectors. The motivation behind the method is to approximate the ideal bootstrap estimates (which corresponds to $B = \infty$) faster than with the usual nonparametric bootstrap approach. In this paper we shall investigate whether this method is able to attenuate the numerical difficulties associated with \mathcal{BBS} log-likelihood function maximization.

3.3. Log-likelihood function with a modified Jeffreys prior penalization. Monotone likelihood cases can arise with considerable frequency in models based on the asymmetric normal distribution, with some samples leading to situations where maximum likelihood estimates of the asymmetry parameter may not be finite, as noted by Liseo (1990). A solution to such problem was proposed by Sartori (2006), who used the score function transformation proposed by Firth (1993) in the asymmetric normal and Student- t models. A more general solution was proposed by Azzalini and Arellano-Valle (2013), who penalized the log-likelihood function as follows:

$$\ell^*(\theta) = \ell(\theta) - Q,$$

where $\ell(\theta)$ and $\ell^*(\theta)$ denote the log-likelihood function and its modified version, respectively. The authors imposed some restrictions on Q , namely: (i) $Q \geq 0$; (ii) $Q = 0$ when the asymmetry parameter equals zero (values close to zero can lead to monotone likelihood cases in the asymmetric normal model); (iii) $Q \rightarrow \infty$ when the asymmetry parameter in absolute value tends to infinity. Additionally, Q should not depend on the data or, at least, be $O_p(1)$. According to Azzalini and Arellano-Valle (2013), when these conditions are satisfied, the estimators obtained using $\ell^*(\theta)$ are finite and have the same asymptotic properties as standard MLEs, such as consistency and asymptotic normality.

FIGURE 3. Q_α and Q_γ , modified Jeffreys penalization.

We shall now use a similar approach for the \mathcal{BBS} model. In particular, we propose modifying the Jeffreys penalization term so that the new penalization satisfies the conditions listed by Azzalini and Arellano-Valle (2013). Since the numerical problems are mainly associated with α and γ , only terms involving these parameters were used. We then arrive at the following penalization term:

$$Q = Q_\gamma + Q_\alpha = -\frac{1}{2} \log \left\{ \frac{(\gamma - \omega)\omega[3 + \gamma(\gamma - \omega)]}{2} + 1 \right\} + \frac{1}{2} \log(1 + \alpha^2), \quad (10)$$

where, as before, $\omega = \phi(\gamma)/\Phi(-\gamma)$.

We note that $Q_\alpha \geq 0$. Additionally, $Q_\alpha \rightarrow 0$ when $\alpha \rightarrow 0$, and $Q_\alpha \rightarrow \infty$ when $\alpha \rightarrow \infty$. It can be shown that $Q_\gamma \rightarrow \infty$ when $\gamma \rightarrow \infty$ and that $Q_\gamma \rightarrow 0$ when $\gamma \rightarrow -\infty$, such that $Q_\gamma \geq 0$. Figure 3 shows the penalization terms as a function of the corresponding parameters. The quantities Q_γ and Q_α penalize large positive values of γ and α , helping avoid estimates that are unexpectedly large. Therefore, the proposed penalization satisfies the conditions indicated by Azzalini and Arellano-Valle (2013). An advantage of the penalization scheme we propose is that, unlike the Jeffreys penalization, it does not require the computation of $L_{\beta\beta}$. In what follows we shall numerically evaluate the effectiveness of the proposed correction when performing point estimation and we shall also consider the issue of carrying out testing inference on the parameters that index the \mathcal{BBS} model.

3.4. Numerical evaluation. A numerical evaluation of the methods described in this section was performed. We considered different \mathcal{BBS} estimation strategies. In what follows we shall focus on the estimation of the bimodality parameter γ .

The Monte Carlo simulations were carried out using the Ox matrix programming language (Doornik, 2009). Numerical maximizations were performed using the the BFGS quasi-Newton method. We considered alternative nonlinear optimization algorithms such as Newton-Raphson and Fisher's scoring, but they did not outperform the BFGS algorithm. We then decided to employ the BFGS method, which is typically regarded as the best performing method (Mittelhammer et al., 2000, Section 8.13). The results are based on 5,000 Monte Carlo replications for values of γ ranging from -2 to 2 and samples of size $n = 50$. In each replication, maximum likelihood estimates were computed and it was verified whether the nonlinear optimization algorithm converged. At the end of the experiment, the frequency of nonconvergences (proportion of samples for which there was no convergence) was computed for each method (denoted by pnf). Figure 4 shows the proportion of nonconvergences corresponding to the standard MLEs, the MLEs obtained using the better bootstrap ($\text{MLE}_{\text{bbboot}}$) and the MLEs obtained from the log-likelihood function penalized using the Jeffreys prior (MLE_{jp}) and its modified version (MLE_{p}) as a function of γ . Notice that the MLE and the $\text{MLE}_{\text{bbboot}}$ are the worst performers when $\gamma > 0$; they display the largest rates of nonconvergence. The methods based on penalized log-likelihood function display the smallest values of pnf, with slight advantage for MLE_{jp} .

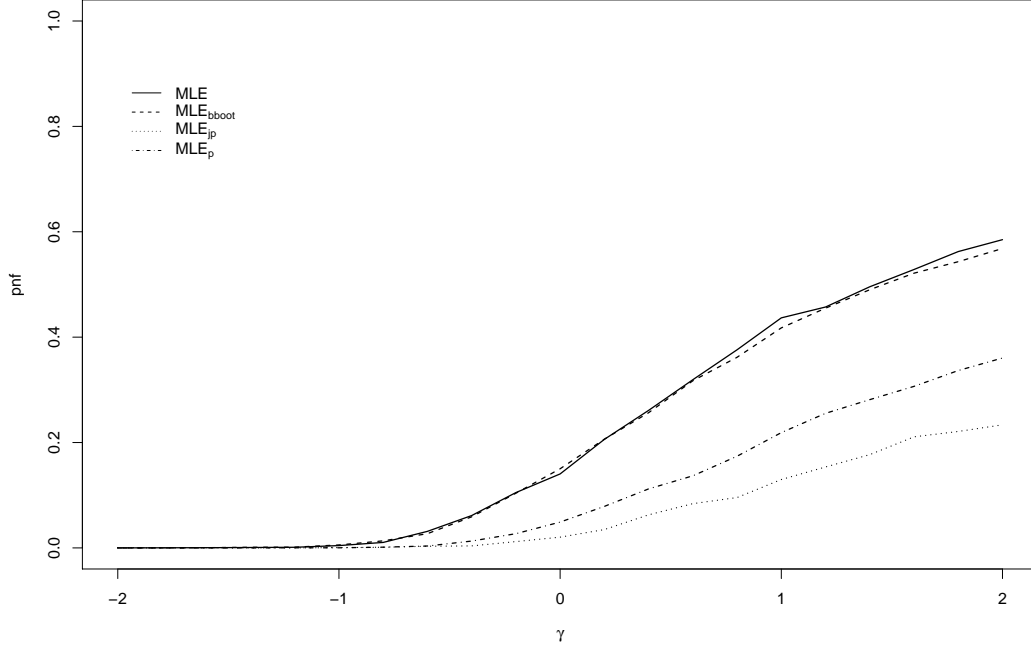
In order to evaluate the impact of the sample size on nonconvergence rates, a numerical study similar to the previous one was performed, but now with the value of the bimodality parameter fixed at $\gamma = 1$. The samples sizes are $n \in \{30, 45, 60, 75, \dots, 300\}$. The number of Monte Carlo replications was 5,000 for each value of n . The results are displayed in Figure 5. We note that the sample size does not seem to influence the MLE and $\text{MLE}_{\text{bbboot}}$ nonconvergence rates. The corresponding optimizations failed in approximately 40% of the samples regardless of the sample size. In contrast, the MLE_{jp} and MLE_{p} failure rates display a slight increase and then stabilize as n increases. Recall that one of the conditions imposed by Azzalini and Arellano-Valle (2013) on the penalization term is that it should remain $O_p(1)$ as $n \rightarrow \infty$, i.e., the penalization influence seems to decrease as larger sample sizes are used, which leads to slightly larger nonconvergence frequencies in larger samples.

A second set of Monte Carlo simulations was carried out, this time only considering the estimator that uses the better bootstrap resampling scheme and also estimators based on the two penalized likelihood functions, i.e., we now only consider $\text{MLE}_{\text{bbboot}}$, MLE_{jp} and MLE_{p} . Again, 5,000 Monte Carlo replications were performed. We estimated the bias (denoted by B) and mean squared errors (denoted by MSE) of the three estimators. The number of nonconvergences is denoted by nf. Tables 1, 2 and 3 contain the results for MLE_{jp} , MLE_{p} and $\text{MLE}_{\text{bbboot}}$, respectively. Overall, MLE_{p} outperforms MLE_{jp} . For instance, when $n = 30$ in the last combination of parameter values, the MSEs of $\hat{\alpha}_{\text{jp}}$, $\hat{\beta}_{\text{jp}}$ and $\hat{\gamma}_{\text{jp}}$ are, respectively, 0.0211, 0.0022 and 2.6671, whereas the corresponding values for $\hat{\alpha}_{\text{p}}$, $\hat{\beta}_{\text{p}}$ and $\hat{\gamma}_{\text{p}}$ are 0.0167, 0.002 and 2.0471. $\text{MLE}_{\text{bbboot}}$ is typically less biased when it comes to the estimation of α and γ , but there are more convergence failures when computing better bootstrap estimates. Overall, the estimator based on the log-likelihood function that uses the penalization term we proposed typically yields more accurate estimates than MLE_{jp} and outperforms $\text{MLE}_{\text{bbboot}}$ in terms of convergence rates.

Next, we shall evaluate how changes in the penalization term impact the frequency of nonconvergences when computing MLE_{p} . In particular, we consider the following penalized log-likelihood function:

$$\ell_{\phi}^*(\theta) = \ell(\theta) - Q^{\phi},$$

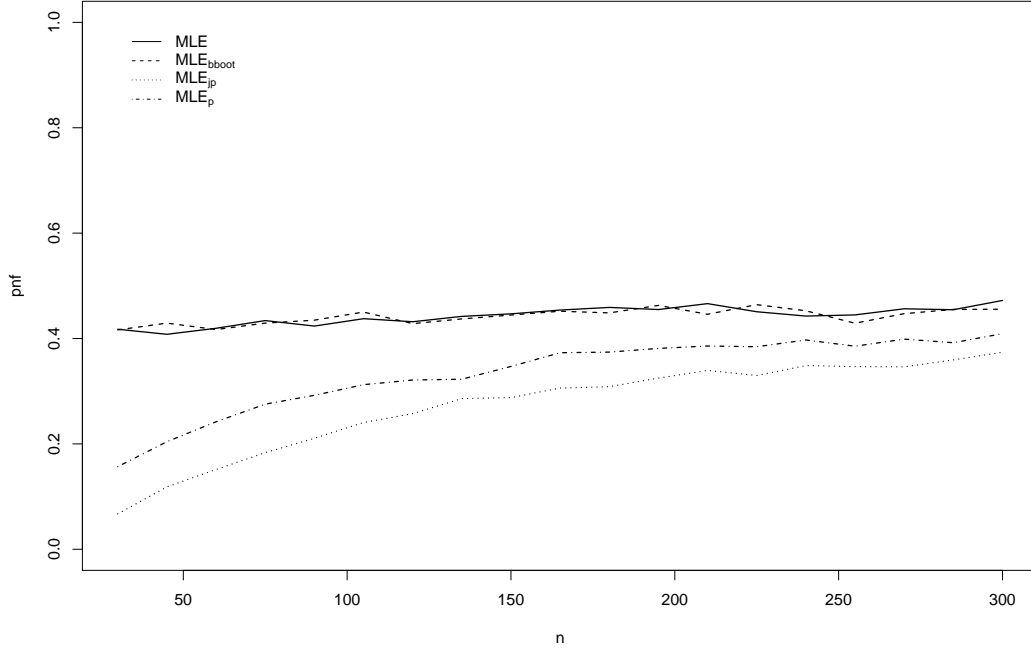
FIGURE 4. Nonconvergence proportions (pnf) for different estimation methods using different values of γ .



with $\phi > 0$ fixed. This additional quantity controls for the penalization strength, with $\phi = 1$ resulting in MLE_ϕ and different values of ϕ leading to stronger or weaker penalizations. A Monte Carlo study was performed to evaluate the accuracy of the parameter estimates for $\gamma \in \{0, 0.1, \dots, 2.0\}$ and $\phi \in \{0.1, 0.2, \dots, 2.1\}$. The parameter values are $\alpha = 0.5$ and $\beta = 1$, and the sample size is $n = 50$. Again, 5,000 replications were performed for each combination of values of ϕ and γ . Samples for which there was convergence failure were discarded. Figure 6 shows the estimated MSEs and the number of non-convergences for each combination of γ and ϕ . Figures 6a and 6b show that estimates of α are less accurate than those of β , both being considerably more accurate than the estimates of γ (Figure 6c). The MSE of the estimator of γ tends to be smaller when the value of ϕ is between 0.4 and 1, especially for larger values of γ . Visual inspection of Figure 6d shows that larger values of γ lead to more non-convergences, which was expected in light of our previous results. Furthermore, nf tends to decrease when larger values of ϕ are used.

We note from Figure 6 that the estimates of γ are the ones most sensitive to changes in the values of γ and ϕ . Figure 7 presents the number of nonconvergences (right vertical axis) and the MSE of $\hat{\gamma}$ (left vertical axis) as a function of γ for three different values of ϕ . Figure 7a shows that, although $\phi = 0.5$ yields more accurate estimates it also leads to more nonconvergences. For $\phi = 1.5$, the number of nonconvergences did not exceed 1400, but $\widehat{\text{MSE}}(\hat{\gamma})$ was larger relative to other values of ϕ . Overall, $\phi = 1$ seems to balance well accuracy and the likelihood of convergence. In what follows we shall use $\phi = 1$.

FIGURE 5. Nonconvergence proportions (pnf) for different estimation methods using different sample sizes.



4. TWO-SIDED HYPOTHESIS TESTS

In this section we consider two-sided hypothesis tests in the \mathcal{BBS} model. Our interest lies in investigating the finite-sample performances of tests based on MLE_p . The first test we consider is the penalized likelihood ratio test, denoted by LR. Consider a model parametrized by $\theta = (\psi, \lambda)$, where ψ is the parameter of interest and λ is a nuisance parameter vector. Our interest lies in testing $H_0 : \psi = \psi_0$ against a two-sided alternative hypothesis. The LR test statistic is

$$W = 2\{\ell^*(\hat{\theta}) - \ell^*(\tilde{\theta})\},$$

where $\hat{\theta}$ is the unrestricted MLE_p of θ , i.e., $\hat{\theta}$ is obtained by maximizing $\ell^*(\theta)$ without imposing restrictions on the parameters, and $\tilde{\theta}$ is the restricted MLE_p , which follows from the maximization of $\ell^*(\theta)$ subject to the restrictions in H_0 . Critical values at the $\epsilon \times 100\%$ significance level are obtained from the null distribution of W which, based on the results in Azzalini and Arellano-Valle (2013), can be approximated by χ_1^2 when ψ is scalar. When ψ is a vector of dimension q (≤ 3), the test is performed in similar fashion with the single difference that the critical value is obtained from χ_q^2 .

It is also possible to test $H_0 : \psi = \psi_0$ against $H_1 : \psi \neq \psi_0$ using the score and Wald tests. To that end, we use the score function and the expected information matrix obtained using the penalized log-likelihood function. The score and Wald test statistics are given, respectively, by

$$W_S = U^*(\tilde{\theta})^\top K^*(\tilde{\theta})^{-1} U^*(\tilde{\theta}),$$

$$W_W = (\hat{\psi} - \psi_0)^2 / K^*(\hat{\theta})^{\psi\psi},$$

FIGURE 6. Mean squared errors of the estimators of α (a), β (b) and γ (c), and the number of nonconvergences nf (d), for different values of γ and ϕ .

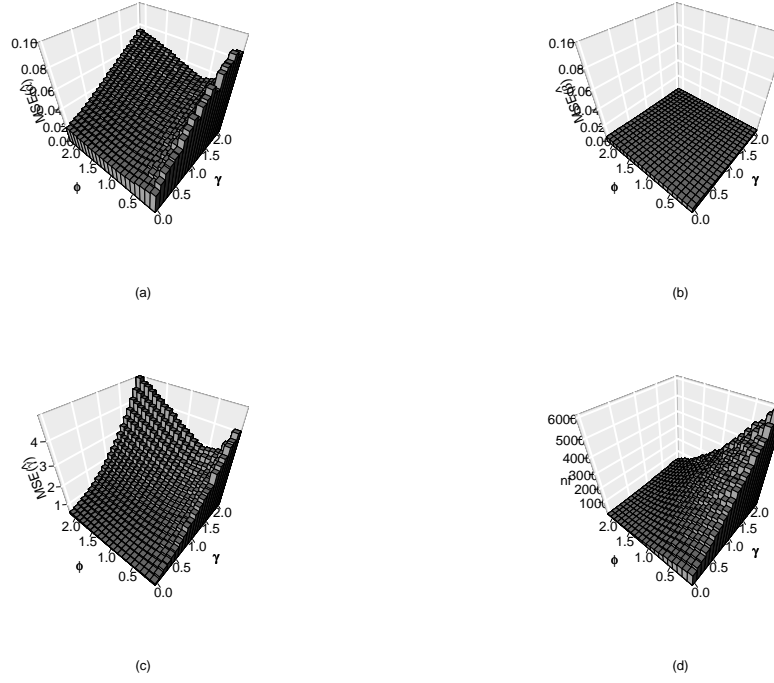


FIGURE 7. Number of nonconvergences (solid line) and MSE of $\hat{\gamma}$ (dashed line) for different values of γ with $\phi \in \{0.5, 1.0, 1.5\}$. The nf values are shown in the left vertical axis and the values of $\widehat{MSE}(\hat{\gamma})$ are shown in the right vertical axis.

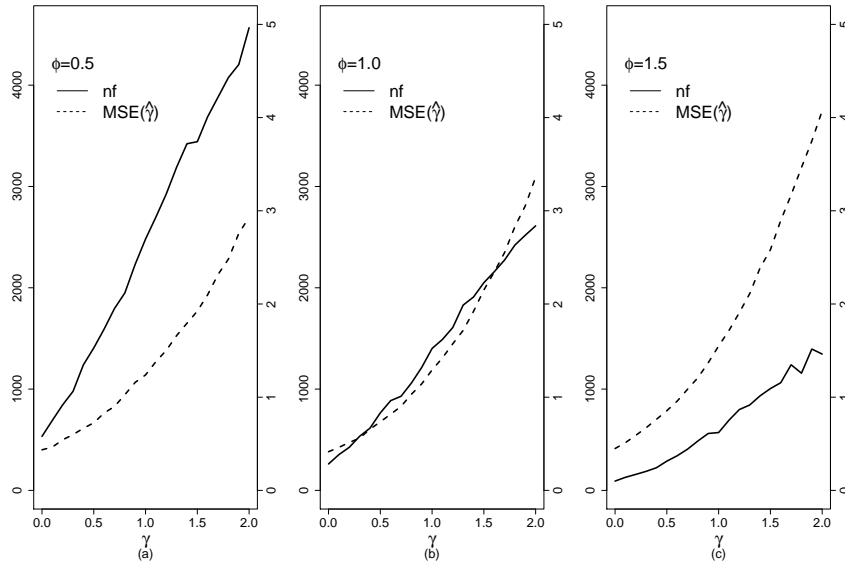


TABLE 1. Bias and mean squared error of MLE_{jp} for some combinations of parameter values.

| n | $\widehat{B}(\hat{\alpha}_{\text{jp}})$ | $\widehat{B}(\hat{\beta}_{\text{jp}})$ | $\widehat{B}(\hat{\gamma}_{\text{jp}})$ | $\widehat{\text{MSE}}(\hat{\alpha}_{\text{jp}})$ | $\widehat{\text{MSE}}(\hat{\beta}_{\text{jp}})$ | $\widehat{\text{MSE}}(\hat{\gamma}_{\text{jp}})$ | nf |
|---|---|--|---|--|---|--|------|
| $\alpha = 0.5, \beta = 1$ and $\gamma = -1$ | | | | | | | |
| 30 | -0.0851 | -0.0048 | -0.4369 | 0.0134 | 0.0118 | 0.3770 | 1 |
| 50 | -0.0567 | -0.0032 | -0.2844 | 0.0079 | 0.0068 | 0.2022 | 7 |
| 100 | -0.0303 | -0.0025 | -0.1498 | 0.0038 | 0.0032 | 0.0876 | 2 |
| 150 | -0.0208 | -0.0016 | -0.1018 | 0.0025 | 0.0021 | 0.0574 | 0 |
| $\alpha = 0.5, \beta = 1$ and $\gamma = 0$ | | | | | | | |
| 30 | -0.1532 | -0.0037 | -0.8487 | 0.0303 | 0.0103 | 0.9668 | 67 |
| 50 | -0.1103 | -0.0017 | -0.5798 | 0.0184 | 0.0057 | 0.5201 | 124 |
| 100 | -0.0692 | -0.0002 | -0.3523 | 0.0095 | 0.0026 | 0.2476 | 185 |
| 150 | -0.0484 | -0.0006 | -0.2412 | 0.0061 | 0.0017 | 0.1510 | 181 |
| $\alpha = 0.5, \beta = 1$ and $\gamma = 1$ | | | | | | | |
| 30 | -0.2295 | 0.0001 | -1.5358 | 0.0585 | 0.0064 | 2.6802 | 373 |
| 50 | -0.1865 | -0.0018 | -1.1795 | 0.0403 | 0.0031 | 1.6360 | 729 |
| 100 | -0.1332 | -0.0009 | -0.8061 | 0.0229 | 0.0012 | 0.8366 | 1520 |
| 150 | -0.1070 | -0.0005 | -0.6337 | 0.0164 | 0.0008 | 0.5793 | 2115 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = -1$ | | | | | | | |
| 30 | -0.0520 | 0.0004 | -0.4429 | 0.0048 | 0.0050 | 0.3812 | 5 |
| 50 | -0.0353 | -0.0005 | -0.2874 | 0.0029 | 0.0028 | 0.2068 | 2 |
| 100 | -0.0174 | -0.0013 | -0.1400 | 0.0013 | 0.0013 | 0.0875 | 2 |
| 150 | -0.0119 | -0.0003 | -0.0970 | 0.0009 | 0.0009 | 0.0578 | 1 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = 0$ | | | | | | | |
| 30 | -0.0914 | -0.0023 | -0.8477 | 0.0109 | 0.0039 | 0.9736 | 86 |
| 50 | -0.0663 | -0.0018 | -0.5873 | 0.0067 | 0.0022 | 0.5309 | 127 |
| 100 | -0.0401 | -0.0003 | -0.3406 | 0.0033 | 0.0010 | 0.2366 | 205 |
| 150 | -0.0298 | 0.0000 | -0.2472 | 0.0022 | 0.0006 | 0.1512 | 208 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = 1$ | | | | | | | |
| 30 | -0.1380 | 0.0002 | -1.5331 | 0.0211 | 0.0022 | 2.6671 | 419 |
| 50 | -0.1114 | 0.0001 | -1.1799 | 0.0144 | 0.0011 | 1.6382 | 745 |
| 100 | -0.0808 | -0.0009 | -0.8149 | 0.0084 | 0.0004 | 0.8538 | 1652 |
| 150 | -0.0649 | 0.0000 | -0.6450 | 0.0059 | 0.0003 | 0.5863 | 2211 |

where $U^*(\theta)$ and $K^*(\theta)$ denote the score function and the expected information, respectively, obtained using the penalized log-likelihood function and $K^*(\theta)^{\psi\psi}$ is the diagonal element of the inverse of $K^*(\theta)$ corresponding to ψ . Both test statistics are asymptotically distributed as χ_1^2 under the null hypothesis.

A Monte Carlo simulation study was performed to evaluate the finite sample performances of the LR, score (denoted by S) and Wald tests in the \mathcal{BBS} model. Log-likelihood maximizations were carried out using the BFGS quasi-Newton method. The number of Monte Carlo replications is 5,000 replications, the sample sizes are $n \in \{30, 50, 100, 150\}$ and the significance levels are $\epsilon \in \{0.1, 0.05, 0.01\}$. The tests were performed for each parameter of the model $\mathcal{BBS}(0.5, 1, 0)$. It is noteworthy that by testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ we test whether the data follows the \mathcal{BS} law, i.e., the original

TABLE 2. Bias and mean squared error of MLE_p for some combinations of parameter values.

| n | $\widehat{B}(\hat{\alpha}_p)$ | $\widehat{B}(\hat{\beta}_p)$ | $\widehat{B}(\hat{\gamma}_p)$ | $\widehat{\text{MSE}}(\hat{\alpha}_p)$ | $\widehat{\text{MSE}}(\hat{\beta}_p)$ | $\widehat{\text{MSE}}(\hat{\gamma}_p)$ | nf |
|---|-------------------------------|------------------------------|-------------------------------|--|---------------------------------------|--|------|
| $\alpha = 0.5, \beta = 1$ and $\gamma = -1$ | | | | | | | |
| 30 | -0.0575 | 0.0078 | -0.3129 | 0.0114 | 0.0121 | 0.3151 | 12 |
| 50 | -0.0364 | 0.0006 | -0.1918 | 0.0072 | 0.0068 | 0.1730 | 4 |
| 100 | -0.0190 | 0.0018 | -0.0999 | 0.0035 | 0.0032 | 0.0795 | 0 |
| 150 | -0.0128 | 0.0004 | -0.0670 | 0.0025 | 0.0021 | 0.0546 | 0 |
| $\alpha = 0.5, \beta = 1$ and $\gamma = 0$ | | | | | | | |
| 30 | -0.1163 | 0.0035 | -0.6441 | 0.0236 | 0.0100 | 0.7143 | 214 |
| 50 | -0.0804 | 0.0042 | -0.4358 | 0.0151 | 0.0056 | 0.4199 | 240 |
| 100 | -0.0477 | 0.0015 | -0.2505 | 0.0082 | 0.0026 | 0.2069 | 287 |
| 150 | -0.0323 | 0.0009 | -0.1697 | 0.0057 | 0.0017 | 0.1369 | 318 |
| $\alpha = 0.5, \beta = 1$ and $\gamma = 1$ | | | | | | | |
| 30 | -0.1965 | 0.0035 | -1.3012 | 0.0474 | 0.0057 | 2.0986 | 857 |
| 50 | -0.1535 | 0.0002 | -0.9703 | 0.0321 | 0.0029 | 1.2726 | 1398 |
| 100 | -0.1058 | 0.0003 | -0.6446 | 0.0185 | 0.0011 | 0.6702 | 2200 |
| 150 | -0.0815 | 0.0011 | -0.4877 | 0.0136 | 0.0007 | 0.4739 | 2688 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = -1$ | | | | | | | |
| 30 | -0.0326 | 0.0012 | -0.2940 | 0.0042 | 0.0049 | 0.3119 | 14 |
| 50 | -0.0197 | 0.0005 | -0.1725 | 0.0025 | 0.0029 | 0.1672 | 7 |
| 100 | -0.0102 | 0.0005 | -0.0909 | 0.0013 | 0.0013 | 0.0846 | 1 |
| 150 | -0.0065 | 0.0002 | -0.0590 | 0.0009 | 0.0009 | 0.0556 | 0 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = 0$ | | | | | | | |
| 30 | -0.0682 | 0.0012 | -0.6436 | 0.0085 | 0.0036 | 0.7375 | 252 |
| 50 | -0.0474 | 0.0020 | -0.4288 | 0.0053 | 0.0021 | 0.4085 | 298 |
| 100 | -0.0277 | 0.0012 | -0.2424 | 0.0029 | 0.0009 | 0.2034 | 349 |
| 150 | -0.0190 | 0.0005 | -0.1686 | 0.0020 | 0.0006 | 0.1340 | 349 |
| $\alpha = 0.3, \beta = 1$ and $\gamma = 1$ | | | | | | | |
| 30 | -0.1161 | 0.0005 | -1.2810 | 0.0167 | 0.0020 | 2.0471 | 979 |
| 50 | -0.0920 | -0.0005 | -0.9769 | 0.0116 | 0.0010 | 1.2924 | 1475 |
| 100 | -0.0624 | 0.0000 | -0.6381 | 0.0067 | 0.0004 | 0.6761 | 2491 |
| 150 | -0.0478 | 0.0005 | -0.4844 | 0.0049 | 0.0003 | 0.4685 | 3066 |

version of the Birnbaum-Saunders distribution. The data were generated according to the model implied by the null hypothesis and samples for which convergence did not take place were discarded.

Tables 4 to 6 contain the null rejection rates of the tests of $H_0 : \alpha = 0.5$, $H_0 : \beta = 1$ and $H_0 : \gamma = 0$, respectively, against two-sided alternative hypotheses. We note that all tests are considerably liberal when the sample size is small (30 or 50). We also note that the score test outperforms the competition. The Wald test was the worst performer.

The tests null rejection rates converge to the corresponding nominal levels as $n \rightarrow \infty$. Such convergence, however, is rather slow. More accurate testing inference can be achieved by using bootstrap resampling; see Davison and Hinkley (1997). The tests employ critical values that are estimated in the bootstrapping scheme instead of asymptotic (approximate) critical values. B bootstrap samples

TABLE 3. Bias and mean squared error of $\text{MLE}_{\text{bboot}}$ for some combinations of parameter values.

| n | $\widehat{B}(\hat{\alpha}_{\text{bboot}})$ | $\widehat{B}(\hat{\beta}_{\text{bboot}})$ | $\widehat{B}(\hat{\gamma}_{\text{bboot}})$ | $\widehat{\text{MSE}}(\hat{\alpha}_{\text{bboot}})$ | $\widehat{\text{MSE}}(\hat{\beta}_{\text{bboot}})$ | $\widehat{\text{MSE}}(\hat{\gamma}_{\text{bboot}})$ | nf |
|--|--|---|--|---|--|---|------|
| $\alpha = 0.5, \beta = 1 \text{ and } \gamma = -1$ | | | | | | | |
| 30 | -0.0283 | 0.0091 | -0.1575 | 0.0141 | 0.0121 | 0.3221 | 126 |
| 50 | -0.0130 | 0.0035 | -0.0761 | 0.0086 | 0.0070 | 0.1898 | 38 |
| 100 | -0.0075 | 0.0017 | -0.0430 | 0.0040 | 0.0033 | 0.0861 | 2 |
| 150 | -0.0041 | 0.0018 | -0.0230 | 0.0027 | 0.0021 | 0.0568 | 0 |
| $\alpha = 0.5, \beta = 1 \text{ and } \gamma = 0$ | | | | | | | |
| 30 | -0.0811 | 0.0056 | -0.4596 | 0.0281 | 0.0093 | 0.7186 | 1494 |
| 50 | -0.0476 | 0.0038 | -0.2658 | 0.0177 | 0.0054 | 0.4071 | 1224 |
| 100 | -0.0215 | 0.0001 | -0.1192 | 0.0093 | 0.0025 | 0.2052 | 860 |
| 150 | -0.0138 | 0.0001 | -0.0780 | 0.0061 | 0.0016 | 0.1365 | 657 |
| $\alpha = 0.5, \beta = 1 \text{ and } \gamma = 1$ | | | | | | | |
| 30 | -0.1609 | 0.0030 | -1.0607 | 0.0498 | 0.0053 | 1.9644 | 4640 |
| 50 | -0.1203 | 0.0026 | -0.7657 | 0.0380 | 0.0028 | 1.2973 | 5096 |
| 100 | -0.0735 | 0.0003 | -0.4606 | 0.0228 | 0.0012 | 0.7213 | 5877 |
| 150 | -0.0562 | 0.0001 | -0.3442 | 0.0154 | 0.0007 | 0.4782 | 5926 |
| $\alpha = 0.3, \beta = 1 \text{ and } \gamma = -1$ | | | | | | | |
| 30 | -0.0159 | 0.0016 | -0.1554 | 0.0050 | 0.0049 | 0.3204 | 150 |
| 50 | -0.0090 | 0.0028 | -0.0820 | 0.0031 | 0.0029 | 0.1927 | 43 |
| 100 | -0.0040 | 0.0008 | -0.0384 | 0.0015 | 0.0013 | 0.0883 | 2 |
| 150 | -0.0032 | 0.0008 | -0.0292 | 0.0009 | 0.0008 | 0.0551 | 0 |
| $\alpha = 0.3, \beta = 1 \text{ and } \gamma = 0$ | | | | | | | |
| 30 | -0.0472 | 0.0024 | -0.4434 | 0.0101 | 0.0034 | 0.7141 | 1543 |
| 50 | -0.0280 | -0.0000 | -0.2633 | 0.0070 | 0.0020 | 0.4315 | 1252 |
| 100 | -0.0127 | 0.0003 | -0.1150 | 0.0034 | 0.0009 | 0.2104 | 813 |
| 150 | -0.0090 | 0.0007 | -0.0823 | 0.0023 | 0.0006 | 0.1416 | 698 |
| $\alpha = 0.3, \beta = 1 \text{ and } \gamma = 1$ | | | | | | | |
| 30 | -0.0983 | 0.0007 | -1.0801 | 0.0179 | 0.0019 | 1.9779 | 4528 |
| 50 | -0.0716 | 0.0002 | -0.7586 | 0.0131 | 0.0010 | 1.2703 | 5128 |
| 100 | -0.0428 | 0.0004 | -0.4492 | 0.0094 | 0.0004 | 0.7788 | 5844 |
| 150 | -0.0328 | -0.0000 | -0.3410 | 0.0059 | 0.0003 | 0.5082 | 6142 |

are generated imposing the null hypothesis and the test statistic is computed for each artificial sample. The critical value of level $\epsilon \times 100\%$ is obtained as the $1 - \epsilon$ upper quantile of the B test statistics, i.e., of the test statistics computed using the bootstrap samples. The bootstrap tests are indicated by the subscript 'pb'. We also use bootstrap resampling to estimate the Bartlett correction factor to the likelihood ratio test as proposed by Rocke (1989). The bootstrap Bartlett corrected test is indicated by the subscript 'bbc'. For details on bootstrap tests, Bartlett-corrected tests and Bartlett corrections based on the bootstrap, the reader is referred to Cordeiro and Cribari-Neto (2014). Since the Wald test proved to be considerably unreliable we shall not consider it.

Next we shall numerically evaluate the finite sample performances of the LR_{pb} , LR_{bbc} and S_{pb} tests under the same scenarios considered for the results presented in Tables 4 to 6. The number of Monte

TABLE 4. Null rejection rates of the LR, score and Wald tests for testing of $H_0 : \alpha = 0.5$ against $H_1 : \alpha \neq 0.5$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

| n | LR | S | Wald |
|-------------------|--------|--------|--------|
| $\epsilon = 0.1$ | | | |
| 30 | 0.2660 | 0.2166 | 0.4460 |
| 50 | 0.2008 | 0.1656 | 0.3374 |
| 100 | 0.1472 | 0.1288 | 0.2392 |
| 150 | 0.1346 | 0.1202 | 0.2078 |
| $\epsilon = 0.05$ | | | |
| 30 | 0.1632 | 0.1116 | 0.3850 |
| 50 | 0.1156 | 0.0820 | 0.2836 |
| 100 | 0.0844 | 0.0664 | 0.1914 |
| 150 | 0.0716 | 0.0574 | 0.1524 |
| $\epsilon = 0.01$ | | | |
| 30 | 0.0508 | 0.0104 | 0.2706 |
| 50 | 0.0352 | 0.0100 | 0.1822 |
| 100 | 0.0230 | 0.0094 | 0.1094 |
| 150 | 0.0166 | 0.0084 | 0.0766 |

TABLE 5. Null rejection rates of the LR, score and Wald tests for testing of $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

| n | LR | S | Wald |
|-------------------|--------|--------|--------|
| $\epsilon = 0.1$ | | | |
| 30 | 0.1692 | 0.0832 | 0.2728 |
| 50 | 0.1416 | 0.0826 | 0.2116 |
| 100 | 0.1190 | 0.0894 | 0.1650 |
| 150 | 0.1170 | 0.0934 | 0.1494 |
| $\epsilon = 0.05$ | | | |
| 30 | 0.1040 | 0.0340 | 0.2162 |
| 50 | 0.0802 | 0.0370 | 0.1480 |
| 100 | 0.0620 | 0.0442 | 0.1090 |
| 150 | 0.0558 | 0.0438 | 0.0808 |
| $\epsilon = 0.01$ | | | |
| 30 | 0.0298 | 0.0048 | 0.1094 |
| 50 | 0.0218 | 0.0080 | 0.0734 |
| 100 | 0.0112 | 0.0076 | 0.0354 |
| 150 | 0.0112 | 0.0090 | 0.0268 |

Carlo replications is as before. Samples for which the optimization methods failed to reach convergence were discard, even for bootstrap samples. The same 1,000 bootstrap samples ($B = 1,000$) were used in all tests. The null rejection rates of the tests for making inferences on α , β and γ are presented in Tables 7, 8 and 9, respectively. It is noteworthy that the tests size distortions are now considerably smaller. For instance, when making inference on α based on a sample of size $n = 30$ and $\epsilon = 0.05$,

TABLE 6. Null rejection rates of the LR, score and Wald tests for testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

| n | LR | S | Wald |
|-------------------|--------|--------|--------|
| $\epsilon = 0.1$ | | | |
| 30 | 0.2570 | 0.2434 | 0.3704 |
| 50 | 0.1974 | 0.1920 | 0.2896 |
| 100 | 0.1368 | 0.1320 | 0.2014 |
| 150 | 0.1364 | 0.1326 | 0.1774 |
| $\epsilon = 0.05$ | | | |
| 30 | 0.1678 | 0.1522 | 0.2992 |
| 50 | 0.1208 | 0.1076 | 0.2134 |
| 100 | 0.0860 | 0.0762 | 0.1454 |
| 150 | 0.0734 | 0.0660 | 0.1140 |
| $\epsilon = 0.01$ | | | |
| 30 | 0.0520 | 0.0334 | 0.1632 |
| 50 | 0.0336 | 0.0234 | 0.1100 |
| 100 | 0.0218 | 0.0150 | 0.0658 |
| 150 | 0.0154 | 0.0114 | 0.0474 |

TABLE 7. Null rejection rates of the bootstrap versions of the LR and score tests of $H_0 : \alpha = 0.5$ against $H_1 : \alpha \neq 0.5$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

| ϵ | LR _{pb} | LR _{bbc} | S _{pb} |
|------------|------------------|-------------------|-----------------|
| $n = 30$ | | | |
| 0.10 | 0.1010 | 0.0934 | 0.0994 |
| 0.05 | 0.0542 | 0.0388 | 0.0500 |
| 0.01 | 0.0100 | 0.0038 | 0.0082 |
| $n = 50$ | | | |
| 0.10 | 0.0990 | 0.0970 | 0.0964 |
| 0.05 | 0.0484 | 0.0420 | 0.0466 |
| 0.01 | 0.0096 | 0.0058 | 0.0108 |
| $n = 100$ | | | |
| 0.10 | 0.1006 | 0.1026 | 0.0988 |
| 0.05 | 0.0510 | 0.0520 | 0.0488 |
| 0.01 | 0.0106 | 0.0102 | 0.0106 |
| $n = 150$ | | | |
| 0.10 | 0.0950 | 0.0970 | 0.0940 |
| 0.05 | 0.0474 | 0.0484 | 0.0486 |
| 0.01 | 0.0096 | 0.0096 | 0.0110 |

the LR and score null rejection rates are 16.32% and 11.16% (Table 4), whereas the corresponding figures for their bootstrap versions LR_{bp}, LR_{bbc} and S_{bp} are 5.42%, 3.88% and 5%, respectively, which are much more closer to 5%. When testing restrictions on β with $n = 30$ and $\epsilon = 0.10$, the LR and score null rejection rates are, respectively, 16.92% and 8.32% (Table 5) whereas their bootstrap versions, LR_{bp}, LR_{bbc} and S_{bp}, display null rejection rates of 10.88%, 10.88% and 11.04%, respectively.

TABLE 8. Null rejection rates of the bootstrap versions of the LR and score tests of $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

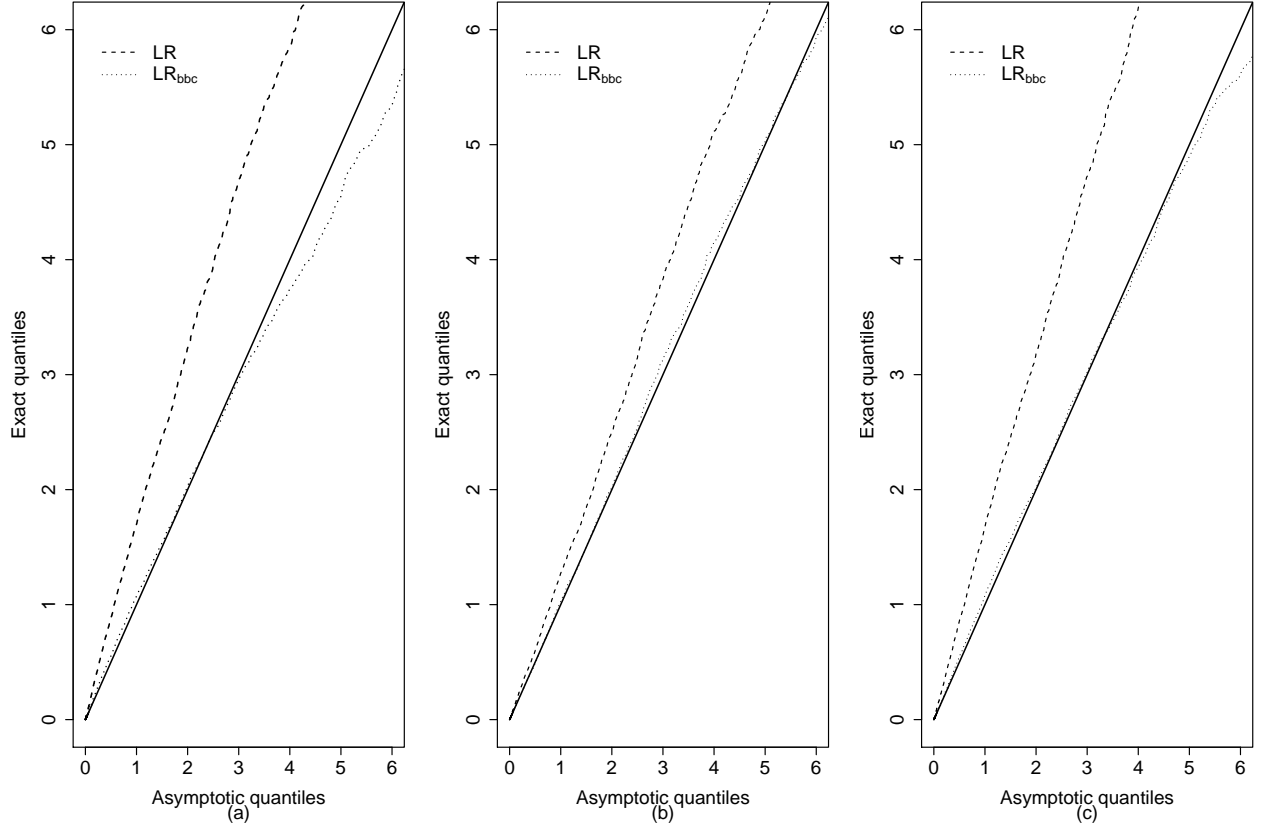
| ϵ | LR _{pb} | LR _{bbc} | S _{pb} |
|------------|------------------|-------------------|-----------------|
| $n = 30$ | | | |
| 0.10 | 0.1088 | 0.1088 | 0.1104 |
| 0.05 | 0.0496 | 0.0490 | 0.0514 |
| 0.01 | 0.0108 | 0.0100 | 0.0114 |
| $n = 50$ | | | |
| 0.10 | 0.1064 | 0.1054 | 0.1096 |
| 0.05 | 0.0532 | 0.0528 | 0.0538 |
| 0.01 | 0.0092 | 0.0094 | 0.0112 |
| $n = 100$ | | | |
| 0.10 | 0.1016 | 0.1002 | 0.1164 |
| 0.05 | 0.0538 | 0.0524 | 0.0574 |
| 0.01 | 0.0102 | 0.0112 | 0.0102 |
| $n = 150$ | | | |
| 0.10 | 0.1048 | 0.1058 | 0.1140 |
| 0.05 | 0.0510 | 0.0510 | 0.0554 |
| 0.01 | 0.0098 | 0.0106 | 0.0104 |

TABLE 9. Null rejection rates of the bootstrap versions of the LR and score tests of $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ in the $\mathcal{BBS}(0.5, 1, 0)$ model.

| ϵ | LR _{pb} | LR _{bbc} | S _{pb} |
|------------|------------------|-------------------|-----------------|
| $n = 30$ | | | |
| 0.10 | 0.1034 | 0.0948 | 0.1022 |
| 0.05 | 0.0522 | 0.0392 | 0.0492 |
| 0.01 | 0.0094 | 0.0030 | 0.0106 |
| $n = 50$ | | | |
| 0.10 | 0.1040 | 0.1028 | 0.1014 |
| 0.05 | 0.0532 | 0.0488 | 0.0514 |
| 0.01 | 0.0098 | 0.0062 | 0.0106 |
| $n = 100$ | | | |
| 0.10 | 0.0980 | 0.1012 | 0.0976 |
| 0.05 | 0.0464 | 0.0488 | 0.0458 |
| 0.01 | 0.0082 | 0.0078 | 0.0088 |
| $n = 150$ | | | |
| 0.10 | 0.0986 | 0.0992 | 0.1014 |
| 0.05 | 0.0468 | 0.0488 | 0.0456 |
| 0.01 | 0.0082 | 0.0080 | 0.0082 |

Finally, when the interest lies in making inferences on γ with $n = 30$ and $\epsilon = 0.01$, the LR and score null rejection rates are 5.20% and 3.34% (Table 6); for the bootstrap-based tests LR_{bp}, LR_{bbc} and S_{bp} we obtain 0.94%, 0.3% and 1.06%, respectively. Figure 8 shows the quantile-quantile (QQ) plots of the

FIGURE 8. Quantile-quantile plots for the LR and LR_{bbc} test statistics with $n = 50$, for the tests on α (a), on β (b) and on γ (c).



LR and LR_{bbc} test statistics for samples of size $n = 50$. It is noteworthy that the empirical quantiles of the LR_{bbc} test statistic are much more closer of the corresponding asymptotic quantiles than those of W . Hence, we note that testing inference in small samples can be made considerably more accurate by using bootstrap resampling.

5. ONE-SIDED HYPOTHESIS TESTS

One-sided tests on a scalar parameter can be performed using the signed likelihood ratio statistic, which is particularly useful in the \mathcal{BBS} model since it allows practitioners to make inferences on γ in a way that makes it possible to detect bimodality. The signed penalized likelihood ratio test statistic is

$$R = \text{sign}(\hat{\psi} - \psi_0) \sqrt{W} = \text{sign}(\hat{\psi} - \psi_0) \sqrt{2\{\ell^*(\hat{\theta}) - \ell^*(\tilde{\theta})\}}. \quad (11)$$

The statistic R is asymptotically distributed as standard normal under the null hypothesis. An advantage of the SLR test over the tests described in Section 4 is that it can be used to perform two-sided and one-sided tests. In this section, we shall focus on one-sided hypothesis testing inference. Our interest lies in detecting bimodality. The null hypothesis is $H_0 : \gamma \geq 0$ which is tested against $H_1 : \gamma < 0$.

Rejection of H_0 yields evidence that the data came from a bimodal distribution. On the other hand, when H_0 is not rejected, there is evidence that the data follows a distribution with a single mode.

Consider the sample $\mathbf{x} = (x_1, \dots, x_n)$ for a model with vector parameter $\theta = (\psi, \lambda)$ with dimension $1 \times p$, where the parameter of interest ψ is a scalar and the vector of nuisance parameters λ has dimension $1 \times (p - 1)$. The test statistic R , given in Equation (11), is asymptotically distributed as standard normal with error of order $O(n^{-1/2})$ when the null hypothesis is true. Such an approximation may not be accurate when the sample size is small. Some analytical corrections for R were proposed in the literature. They can be used to improve the test finite sample behavior.

An important contribution was made by Barndorff-Nielsen (1986, 1991). The author proposed a correction term \mathcal{U} of the form

$$R^* = R + \log(\mathcal{U}/R)/R,$$

where R represents the SLR statistic and R^* is its corrected version. Let $\ell(\theta)$ be the log-likelihood function of the parameters. Its derivatives will be denoted here by

$$\ell_{\theta}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} \quad \text{and} \quad \ell_{\theta\theta}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top}.$$

The observed information matrix is given by $J_{\theta\theta}(\theta) = -\ell_{\theta\theta}(\theta)$. To obtain the correction proposed by Barndorff-Nielsen (1986), the sufficient statistic has to be of the form $(\hat{\theta}, a)$, where $\hat{\theta}$ is the MLE of θ and a is an ancillary statistic. Additionally, it is necessary to compute sample space derivatives of the log-likelihood, such as

$$\ell_{,\hat{\theta}}(\theta) = \frac{\partial \ell(\theta)}{\partial \hat{\theta}} \quad \text{and} \quad \ell_{\theta;\hat{\theta}}(\theta) = \frac{\partial \ell_{,\hat{\theta}}(\theta)}{\partial \theta^\top},$$

where derivatives are taken with respect to some functions of the sample while keeping other terms fixed, as explained in Severini (2000). The quantity \mathcal{U} is given by

$$\mathcal{U} = \frac{\left| \begin{array}{c} \ell_{,\hat{\theta}}(\hat{\theta}) - \ell_{,\hat{\theta}}(\tilde{\theta}) \\ \ell_{\lambda;\hat{\theta}}(\tilde{\theta}) \end{array} \right|}{|J_{\lambda\lambda}(\tilde{\theta})|^{1/2} |J_{\theta\theta}(\hat{\theta})|^{1/2}},$$

where $\tilde{\theta}$ is the restricted MLE of θ and the indices indicate which components are being used in each vector or matrix.

The null distribution of R^* is standard normal with error of order $O(n^{-3/2})$. Although the null distribution of R^* is better approximated by the limiting distribution than that of R , the computation of \mathcal{U} is restricted to some specific classes of models, such as exponential family and transformation models (Severini, 2000).

Some alternatives to R^* were proposed in the literature. They approximate the sample space derivatives used in \mathcal{U} . For instance, approximations were obtained by DiCiccio and Martin (1993), Fraser et al. (1999) and Severini (1999). They were computed by Wu and Wong (2004) for the \mathcal{BS} model and by Lemonte and Ferrari (2011) for a Birnbaum-Saunders regression model. Other recent contributions are Ferrari and Pinheiro (2016) and Smith et al. (2015).

In this paper, we apply the approximations proposed by Severini (1999) and Fraser et al. (1999) using the log-likelihood without penalization. Our interest is to evaluate the performance of the corrections when applied for the statistic R , calculated using the penalized log-likelihood function, comparing the performance of the corrected tests with the SLR and its bootstrap version.

Using the same notation as Lemonte and Ferrari (2011), the approximation proposed by Fraser et al. (1999) (denoted by SLR_{c1}) for \mathcal{U} can be written as

$$\mathcal{U}_1 = \frac{\begin{vmatrix} \Gamma_\theta \\ \Psi_{\lambda\theta} \end{vmatrix}}{|J_{\lambda\lambda}(\tilde{\theta})|^{1/2} |J_{\theta\theta}(\hat{\theta})|^{1/2}},$$

where

$$\Gamma_\theta = [\ell_{\cdot;\mathbf{x}}(\hat{\theta}) - \ell_{\cdot;\mathbf{x}}(\tilde{\theta})] V(\hat{\theta}) [\ell_{\theta;\mathbf{x}}(\hat{\theta}) V(\hat{\theta})]^{-1} J_{\theta\theta}(\hat{\theta}),$$

with

$$\Psi_{\theta\theta} = \begin{bmatrix} \Psi_{\psi\theta} \\ \Psi_{\lambda\theta} \end{bmatrix} = \ell_{\theta;\mathbf{x}}(\tilde{\theta}) V(\hat{\theta}) [\ell_{\theta;\mathbf{x}}(\hat{\theta}) V(\hat{\theta})]^{-1} J_{\theta\theta}(\hat{\theta}),$$

where $\ell_{\cdot;\mathbf{x}}(\theta) = \partial l(\theta) / \partial \mathbf{x}$ is a $1 \times n$ vector, $\ell_{\theta;\mathbf{x}}(\tilde{\theta}) = \partial^2 \ell(\theta) / \partial \theta^\top \partial \mathbf{x}$ is a $p \times n$ matrix and

$$V(\theta) = - \left[\frac{\partial \mathbf{z}(\mathbf{x}; \theta)}{\partial \mathbf{x}} \right]^{-1} \left[\frac{\partial \mathbf{z}(\mathbf{x}; \theta)}{\partial \theta^\top} \right]$$

is an $n \times p$ matrix, $\mathbf{z}(\mathbf{x}; \theta)$ being a vector of pivotal quantities.

We used the distribution function of $Y = |T| + \gamma$ to obtain SLR_{c1} in the \mathcal{BBS} model, where Y follows the normal distribution truncated in (γ, ∞) . The corrected SLR statistic obtained using the approximation given by Fraser et al. (1999) is $R_{c1} = R + \log(\mathcal{U}_1 / R) / R$, which has asymptotic standard normal distribution with error of order $O(n^{-3/2})$ under the null hypothesis. We computed the quantities needed to obtain \mathcal{U}_1 in the \mathcal{BBS} model, which are presented below.

Consider the variable $Y = |T| + \gamma$, where $T = \alpha^{-1}(\sqrt{X/\beta} - \sqrt{\beta/X})$ and $X \sim \mathcal{BBS}(\alpha, \beta, \gamma)$. The distribution of Y is truncated standard normal distribution with support (γ, ∞) , its distribution function being given by

$$F_Y(y) = \begin{cases} 0 & \text{if } y < \gamma, \\ \frac{\Phi(y) - \Phi(\gamma)}{1 - \Phi(\gamma)} & \text{if } y \geq \gamma. \end{cases}$$

Therefore, $Z = F_Y(Y)$ is uniformly distributed in the standard interval, $(0, 1)$. Hence, it is a pivotal quantity that can be used for obtaining the approximations to sample space derivatives proposed by Fraser et al. (1999). Let $\mathbf{x} = (x_1, \dots, x_n)$ be a random $\mathcal{BBS}(\alpha, \beta, \gamma)$ sample. It follows that $\partial z_i / \partial x_j = 0$ when $i \neq j$ and $\partial z_i / \partial x_i = \phi(y_i) \text{sign}(t_i)(x_i + \beta) / [\Phi(-\gamma) 2\alpha\beta^{1/2} x_i^{3/2}]$, with $t_i = \alpha^{-1}(\sqrt{x_i/\beta} - \sqrt{\beta/x_i})$. Moreover, $\partial z_i / \partial \alpha = -\phi(y_i) \text{sign}(t_i) t_i / \Phi(-\gamma) \alpha$, $\partial z_i / \partial \beta = -\phi(y_i) \text{sign}(t_i) [\sqrt{x_i/\beta} + \sqrt{\beta/x_i}] / [\Phi(-\gamma) 2\alpha\beta]$ and $\partial z_i / \partial \gamma = [\Phi(y_i) - \Phi(\gamma)] \phi(\gamma) / \Phi^2(-\gamma) + [\phi(y_i) - \phi(\gamma)] / \Phi(-\gamma)$, where $y_i = |t_i| + \gamma$ and $z_i = F_Y(y_i)$. Therefore, $v_{\alpha i} = 2\beta^{1/2} x_i^{3/2} t_i / (x_i + \beta)$, $v_{\beta i} = x_i / \beta$ and $v_{\gamma i} = -2\alpha\beta^{1/2} x_i^{3/2} \{[\Phi(y_i) - \Phi(\gamma)] \phi(\gamma) / \Phi(-\gamma) + \phi(y_i) - \phi(\gamma)\} / [\phi(y_i) \text{sign}(t_i)(x_i + \beta)]$. The vectors \mathbf{v}_α , \mathbf{v}_β and \mathbf{v}_γ are used to form the matrix $V(\theta)$. For instance, $\mathbf{v}_\alpha = (v_{\alpha 1}, \dots, v_{\alpha n})^\top$ is a $n \times 1$ vector. Here, $V(\theta) = [\mathbf{v}_\alpha \mathbf{v}_\beta \mathbf{v}_\gamma]$. Furthermore, we have that

$$\begin{aligned} \ell_{\cdot; x_i}(\theta) &= \frac{-3}{2x_i} + \frac{1}{x_i + \beta} - (|t_i| + \gamma) \text{sign}(t_i) \frac{(x_i + \beta)}{2\alpha\beta^{1/2} x_i^{3/2}}, \\ \ell_{\alpha; x_i}(\theta) &= \frac{\text{sign}(t_i)(x_i + \beta)}{2\beta^{1/2} x_i^{3/2} \alpha^2} (2|t_i| + \gamma), \\ \ell_{\beta; x_i}(\theta) &= \frac{(-1)}{(x_i + \beta)^2} + \frac{(x_i + \beta)}{4\alpha^2 \beta^{3/2} x_i^{3/2}} \left(\frac{x_i^{1/2}}{\beta^{1/2}} + \frac{\beta^{1/2}}{x_i^{1/2}} \right) + \frac{\text{sign}(t_i)(|t_i| + \gamma)(x_i - \beta)}{4\alpha\beta^{3/2} x_i^{3/2}}, \end{aligned}$$

$$\ell_{\gamma; x_i}(\theta) = -\frac{\text{sign}(t_i)(x_i + \beta)}{2\alpha\beta^{1/2}x_i^{3/2}}.$$

The method proposed by Severini (1999) (denoted by SLR_{c2}) approximates the sample space derivatives by covariances of the log-likelihood function. The main idea is to use the sample to obtain the covariance values empirically. Using again the notation of Lemonte and Ferrari (2011), the approximation of \mathcal{U} proposed by Severini (1999) is given by

$$\mathcal{U}_2 = \frac{\left| \begin{array}{c} \Delta_\theta \\ \Sigma_{\lambda\theta} \end{array} \right|}{|J_{\lambda\lambda}(\tilde{\theta})|^{1/2} |J_{\theta\theta}(\hat{\theta})|^{1/2}},$$

with

$$\Delta_\theta = [Q(\hat{\theta}; \hat{\theta}) - Q(\tilde{\theta}; \hat{\theta})] I(\hat{\theta}; \hat{\theta})^{-1} J_{\theta\theta}(\hat{\theta})$$

and

$$\Sigma_{\theta\theta} = \begin{bmatrix} \Sigma_{\psi\theta} \\ \Sigma_{\lambda\theta} \end{bmatrix} = I(\tilde{\theta}; \hat{\theta}) I(\hat{\theta}; \hat{\theta})^{-1} J_{\theta\theta}(\hat{\theta}),$$

where $Q(\theta; \theta_0) = \sum_{i=1}^n \ell^{(i)}(\theta) \ell_\theta^{(i)}(\theta_0)^\top$ is a $1 \times p$ vector and $I(\theta; \theta_0) = \sum_{i=1}^n \ell_\theta^{(i)}(\theta) \ell_\theta^{(i)}(\theta_0)^\top$ is a $p \times p$ matrix, the index (i) indicating that the quantity corresponds to the i th sample observation. The corrected statistic proposed by Severini (1999) is $R_{c2} = R + \log(\mathcal{U}_2/R)/R$. Its null distribution is standard normal with error of order $O(n^{-1})$. The score function and the observed information matrix, which can be found in Olmos et al. (2016), are used to obtain \mathcal{U}_2 in the \mathcal{BBS} model.

Alternatively, bootstrapping resampling can be used to obtain critical values for the SLR test. Since we test $H_0 : \gamma \geq 0$ against $H_1 : \gamma < 0$, the critical value of level $\epsilon \times 100\%$ is obtained as the ϵ quantile of the B test statistics computed using the bootstrap samples.

A simulation study was performed to evaluate the sizes and powers of the SLR, SLR_{c1} , SLR_{c2} and SLR_{bp} tests. We tested $H_0 : \gamma \geq 0$ against $H_1 : \gamma < 0$. The true parameter values are $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$. The most reliable tests are those with large power (i.e., higher probability of rejecting H_0 when $\gamma < 0$) and small size distortions. Again, 5,000 Monte Carlo replication were performed. The SLR_{bp} test is based on 1,000 bootstrap samples. The simulation results are presented in Table 10. The most powerful tests are SLR, SLR_{c2} and SLR_{c1} , in that order, whereas the tests with the smallest size distortions are SLR_{bp} , SLR_{c1} and SLR_{c2} . We recommend that testing inference be based on either SLR_{c1} or SLR_{c2} , since these tests display a good balance between size and power.

6. NONNESTED HYPOTHESIS TESTS FOR THE BIMODAL BIRNBAUM-SAUNDERS MODEL

In the previous section we presented a test that is useful for detecting whether the data came from a bimodal \mathcal{BBS} law. That was done by testing a restriction on γ . In this section we shall present tests that are useful for distinguishing between \mathcal{BBS} model and another extension of the \mathcal{BS} distribution that can display bimodality.

As noted in the Introduction, another variation of the \mathcal{BS} distribution that can exhibit bimodality is the model recently discussed by Owen and Ng (2015), which the authors denoted by \mathcal{GBS}_2 . Let $X \sim \mathcal{GBS}_2(\alpha, \beta, \nu)$. Its probability density function is given by

$$g(x) = \frac{\nu}{\alpha x} \left[\left(\frac{x}{\beta} \right)^\nu + \left(\frac{\beta}{x} \right)^\nu \right] \phi \left(\frac{1}{\alpha} \left[\left(\frac{x}{\beta} \right)^\nu - \left(\frac{\beta}{x} \right)^\nu \right] \right), \quad x > 0,$$

TABLE 10. Null rejection rates of the SLR, SLR_{c1} , SLR_{c2} and SLR_{bp} tests of $H_0 : \gamma \geq 0$ against $H_1 : \gamma < 0$ in a sample of size 30 of the model $\mathcal{BBS}(0.5, 1, \gamma)$.

| ϵ | SLR | SLR_{c1} | SLR_{c2} | SLR_{bp} |
|-----------------|--------|-------------------|-------------------|-------------------|
| $\gamma = -1$ | | | | |
| 0.10 | 0.7488 | 0.5768 | 0.6310 | 0.4960 |
| 0.05 | 0.6300 | 0.4276 | 0.4728 | 0.3488 |
| 0.01 | 0.3560 | 0.1910 | 0.2100 | 0.1376 |
| $\gamma = -0.5$ | | | | |
| 0.10 | 0.4634 | 0.2766 | 0.3300 | 0.2248 |
| 0.05 | 0.3242 | 0.1762 | 0.2042 | 0.1326 |
| 0.01 | 0.1328 | 0.0588 | 0.0652 | 0.0376 |
| $\gamma = 0$ | | | | |
| 0.10 | 0.2614 | 0.1334 | 0.1678 | 0.1042 |
| 0.05 | 0.1658 | 0.0746 | 0.0892 | 0.0498 |
| 0.01 | 0.0486 | 0.0202 | 0.0210 | 0.0106 |
| $\gamma = 0.5$ | | | | |
| 0.10 | 0.1546 | 0.0722 | 0.0928 | 0.0526 |
| 0.05 | 0.0890 | 0.0378 | 0.0442 | 0.0222 |
| 0.01 | 0.0214 | 0.0098 | 0.0102 | 0.0046 |
| $\gamma = 1$ | | | | |
| 0.10 | 0.1144 | 0.0488 | 0.0640 | 0.0340 |
| 0.05 | 0.0606 | 0.0246 | 0.0292 | 0.0150 |
| 0.01 | 0.0144 | 0.0050 | 0.0046 | 0.0022 |

where $\alpha > 0$, $\beta > 0$ and $\nu > 0$. According to Owen and Ng (2015), the \mathcal{GBS}_2 density is bimodal when $\alpha > 2$ and $\nu > 2$ (simultaneously).

Therefore, when bimodality is detected the data analysis may be carried out with either the \mathcal{BBS} distribution or the \mathcal{GBS}_2 model. It would then be useful to have a hypothesis test that could be used to distinguish between the two models. Obviously, the tests discussed so far cannot be used to that end. \mathcal{BS} model selection criteria were considered by Leiva (2015) and Leiva et al. (2015). Model selection is usually based on the Bayes factor and also on the Schwarz and Akaike information criteria. We shall use a different approach: we shall develop tests for nonnested hypotheses. Notice that the \mathcal{GBS}_2 distribution cannot be obtained from the \mathcal{BBS} distribution by imposing restrictions on the model parameters, and vice-versa. Hence, the two models are not nested.

The literature of nonnested models began with Cox (1961, 1962). The author introduced likelihood ratio tests for some nonnested models. His main results were generalized by Vuong (1989), who considered nested, nonnested and overlapping models and derived the required asymptotics. For nonnested models, Vuong (1989) established the relationship between the likelihood ratio statistic and the Kullback-Leibler information. Let F and G be competing nonnested models. The author presented a test of the null hypothesis H_0 that both models are equivalent, the alternative hypotheses being: H_f : model F is better and H_g : model G is better. An alternative approach for testing of nonnested models was considered by Williams (1970) and Lewis et al. (2011). The authors only considered tests of the hypothesis H_f and H_g . They proposed to consider H_f and H_g sequentially.

We shall consider the hypothesis involving the \mathcal{BBS} and \mathcal{GBS}_2 models as:

- H_f - the data came from the \mathcal{BBS} distribution,
- H_g - the data came from the \mathcal{GBS}_2 distribution.

The test statistic we considered is the following likelihood ratio statistic:

$$W_{ne} = \log \left(\frac{\hat{f}}{\hat{g}} \right) = \hat{\ell}_f - \hat{\ell}_g,$$

where \hat{f} and \hat{g} denote the likelihood functions of the \mathcal{BBS} and \mathcal{GBS}_2 models, respectively, evaluated at the respective MLE estimates, ℓ representing the log-likelihood function of the model indicated by its index. Then, for a given sample \mathbf{x} , a large positive value of W_{ne} yields evidence in favor H_f and against H_g ; on the other hand, a large negative value of W_{ne} favors H_g . The \mathcal{BBS} parameters are estimated using the penalized log-likelihood function and those of \mathcal{GBS}_2 are estimated using the standard log-likelihood function.

In the test discussed by Vuong (1989) for nonnested models, the asymptotic null distribution is standard normal. On some Monte Carlo simulations not reported here, this test, based on asymptotic critical values, was observed to indicate equivalence of the models at a large rate. Since this test is based on large sample approximations, a bootstrap resampling method could be used as an alternative to provide more accurate critical values and to indicate only one of the models with higher frequency. On the other hand, application of bootstrap resampling is not straightforward for this kind of test because we would have to define an model equivalent to \mathcal{BBS} and \mathcal{GBS}_2 to generate pseudo-samples under the null distribution. Thus, an approach similar to the one taken by Lewis et al. (2011) will be used in this paper, considering only the hypothesis H_f and H_g in the test. Then, taking H_f as null hypothesis, the bootstrap test is done using the following procedures:

- (1) Calculate the value of W_{ne} for the sample \mathbf{x} ;
- (2) With the MLE estimates of the parameters from the \mathcal{BBS} model, generate a bootstrap sample \mathbf{x}^* , and then compute W_{ne}^* for this sample;
- (3) Execute step 2 for B times and obtain the p -value bootstrap: $p_b = \frac{\#\{W_{ne}^* < W_{ne}\} + 1}{B + 1}$.

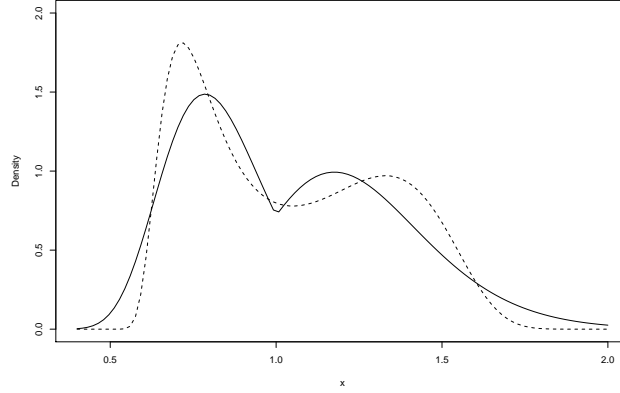
Hence, at the significance level of $\epsilon \times 100\%$, H_f is rejected if $p_b < \epsilon$, i.e., we reject that the data was originated from a \mathcal{BBS} and accept that the \mathcal{GBS}_2 distribution is more adequate. The same can be done taking H_g as null hypothesis. In this case, the procedures are the following:

- (1) Calculate the value of W_{ne} for the sample \mathbf{x} ;
- (2) With the MLE estimates of the parameters from the \mathcal{GBS}_2 model, generate a bootstrap sample \mathbf{x}^* , and then compute W_{ne}^* for this sample;
- (3) Execute step 2 for B times and obtain the p -value bootstrap: $p_b = \frac{\#\{W_{ne}^* > W_{ne}\} + 1}{B + 1}$.

It is noteworthy that the step 3 is different from the corresponding step in the first procedure, since the rejection region changes when we consider H_g as null hypothesis. Again, at the significance level of $\epsilon \times 100\%$, the null hypothesis considered is rejected if $p_b < \epsilon$, but in this case it means that we reject the hypothesis H_g that the data comes from a \mathcal{GBS}_2 distribution and accept that \mathcal{BBS} model is more adequate.

The problem with this approach is that four inferences results are possible to happen, as cited by Williams (1970) and Lewis et al. (2011):

- R1 Both hypothesis H_f and H_g are not rejected, indicating that both models are adequate;
- R2 We do not reject H_f , but H_g is rejected, indicating that the \mathcal{BBS} model is more adequate;
- R3 We do not reject H_g , but H_f is rejected, indicating that the \mathcal{GBS}_2 model is more adequate;
- R4 We reject both H_f and H_g , indicating that both models are not adequate;

FIGURE 9. Densities $\mathcal{BB}\mathcal{S}(0.2, 1, -1)$ (solid line) and $\mathcal{GB}\mathcal{S}_2(5, 1, 5)$ (dashed line).

Under some regularity conditions, Vuong (1989) has shown that, in nonnested models, an adjusted likelihood ratio tends to infinity under H_f when $n \rightarrow \infty$ and that under H_g it tends to minus infinity when $n \rightarrow \infty$. This way, the test statistic tends to indicate the correct model as the sample size increases. Therefore, for a situation where we obtain the result $R1$, a way to choose between the two models is to use the statistic W_{ne} calculated for the sample, opting for the $\mathcal{BB}\mathcal{S}$ distribution in case $W_{ne} > 0$ and for the $\mathcal{GB}\mathcal{S}_2$ distribution in case $W_{ne} < 0$.

TABLE 11. Proportions of outcomes of the test of H_f against H_g when the data generating function is $\mathcal{BB}\mathcal{S}(0.2, 1, -1)$ (first four columns), and proportions of $\mathcal{BB}\mathcal{S}$ and $\mathcal{GB}\mathcal{S}_2$ model selection and porportion of no model selected. The test significance level is ϵ .

| ϵ | $R1$ | $R2$ | $R3$ | $R4$ | $\mathcal{BB}\mathcal{S}$ | $\mathcal{GB}\mathcal{S}_2$ | None |
|------------|--------|--------|--------|--------|---------------------------|-----------------------------|--------|
| $n = 30$ | | | | | | | |
| 0.10 | 0.6676 | 0.2378 | 0.0946 | 0.0000 | 0.4354 | 0.5646 | 0.0000 |
| 0.05 | 0.8284 | 0.1322 | 0.0394 | 0.0000 | 0.4354 | 0.5646 | 0.0000 |
| 0.01 | 0.9650 | 0.0292 | 0.0058 | 0.0000 | 0.4354 | 0.5646 | 0.0000 |
| $n = 50$ | | | | | | | |
| 0.10 | 0.5454 | 0.3578 | 0.0968 | 0.0000 | 0.5574 | 0.4426 | 0.0000 |
| 0.05 | 0.7230 | 0.2326 | 0.0444 | 0.0000 | 0.5572 | 0.4428 | 0.0000 |
| 0.01 | 0.9174 | 0.0754 | 0.0072 | 0.0000 | 0.5572 | 0.4428 | 0.0000 |
| $n = 100$ | | | | | | | |
| 0.10 | 0.3420 | 0.5538 | 0.1036 | 0.0006 | 0.6952 | 0.3042 | 0.0006 |
| 0.05 | 0.5344 | 0.4138 | 0.0518 | 0.0000 | 0.6940 | 0.3060 | 0.0000 |
| 0.01 | 0.8024 | 0.1884 | 0.0092 | 0.0000 | 0.6940 | 0.3060 | 0.0000 |
| $n = 150$ | | | | | | | |
| 0.10 | 0.1920 | 0.6908 | 0.1138 | 0.0034 | 0.7682 | 0.2284 | 0.0034 |
| 0.05 | 0.3772 | 0.5646 | 0.0582 | 0.0000 | 0.7656 | 0.2344 | 0.0000 |
| 0.01 | 0.6720 | 0.3118 | 0.0162 | 0.0000 | 0.7654 | 0.2346 | 0.0000 |

TABLE 12. Proportions of outcomes of the test of H_f against H_g when the data generating function is $\mathcal{GBS}_2(5, 1, 5)$ (first four columns), and proportions of \mathcal{BBS} and \mathcal{GBS}_2 model selection and porportion of no model selected. The test significance level is ϵ .

| ϵ | $R1$ | $R2$ | $R3$ | $R4$ | \mathcal{BBS} | \mathcal{GBS}_2 | None |
|------------|--------|--------|--------|--------|-----------------|-------------------|--------|
| $n = 30$ | | | | | | | |
| 0.10 | 0.4438 | 0.1080 | 0.4482 | 0.0000 | 0.1784 | 0.8216 | 0.0000 |
| 0.05 | 0.6658 | 0.0538 | 0.2804 | 0.0000 | 0.1784 | 0.8216 | 0.0000 |
| 0.01 | 0.9040 | 0.0112 | 0.0848 | 0.0000 | 0.1784 | 0.8216 | 0.0000 |
| $n = 50$ | | | | | | | |
| 0.10 | 0.2204 | 0.1010 | 0.6786 | 0.0000 | 0.1270 | 0.8730 | 0.0000 |
| 0.05 | 0.4502 | 0.0466 | 0.5032 | 0.0000 | 0.1238 | 0.8762 | 0.0000 |
| 0.01 | 0.7624 | 0.0090 | 0.2286 | 0.0000 | 0.1238 | 0.8762 | 0.0000 |
| $n = 100$ | | | | | | | |
| 0.10 | 0.0146 | 0.0662 | 0.8804 | 0.0388 | 0.0672 | 0.8940 | 0.0388 |
| 0.05 | 0.1068 | 0.0560 | 0.8350 | 0.0022 | 0.0712 | 0.9266 | 0.0022 |
| 0.01 | 0.4000 | 0.0112 | 0.5888 | 0.0000 | 0.0674 | 0.9326 | 0.0000 |
| $n = 150$ | | | | | | | |
| 0.10 | 0.0002 | 0.0158 | 0.8968 | 0.0872 | 0.0158 | 0.8970 | 0.0872 |
| 0.05 | 0.0136 | 0.0310 | 0.9344 | 0.0210 | 0.0320 | 0.9470 | 0.0210 |
| 0.01 | 0.1610 | 0.0122 | 0.8268 | 0.0000 | 0.0322 | 0.9678 | 0.0000 |

A simulation study was done to evaluate the performance of the nonnested hypothesis tests involving the \mathcal{BBS} and \mathcal{GBS}_2 distributions. The models considered were $\mathcal{BBS}(0.2, 1, -1)$ and $\mathcal{GBS}_2(5, 1, 5)$. Figure 9 shows their density plots, where we can see they have close density shapes. The number of Monte Carlo replications used was 5000. First, we considered the case when the true distribution is $\mathcal{BBS}(0.2, 1, -1)$; in each replication, $B = 1000$ bootstrap samples were generated under H_f and another $B = 1000$ under H_g , obtaining then, one of the results previously cited ($R1$, $R2$, $R3$ or $R4$) for the test in each replication analyzed. Table 11 contains the proportion of times each result occurred for this simulation and the proportion of time each distribution is chosen as the most suitable model, which is: the \mathcal{BBS} model, when we obtain result R_1 and $W_{ne} > 0$ or we obtain result R_2 ; the \mathcal{GBS}_2 model, when we obtain result R_1 and $W_{ne} < 0$ or we obtain result R_3 ; none of the distributions, when we obtain result R_4 . The same procedures were done when the true model is the $\mathcal{GBS}_2(5, 1, 5)$ distribution, with results presented in Table 12.

From the results contained in Table 11, we can note that the null rejection rates of the true hypothesis (H_f), are close of the nominal levels considered. For instance, when $n = 30$ and $\epsilon = 0.10$, adding the cells corresponding to $R3$ and $R4$, we see that the rejection rate of H_f is 9.46%, a rate close of the nominal level considered. We can also note that for small or moderate sample sizes the tests tend to indicate equivalence of both models, but when the value of n increases, the tests tend to indicate the \mathcal{BBS} model as the most suitable with higher frequency. When $\epsilon = 0.05$, for example, in the column corresponding to $R2$, this happens 13.22% of the time for $n = 30$, while for $n = 150$, in 56.46% of the time the \mathcal{BBS} distribution is considered the most adequate model. This can be observed also in the fifth column, where we can see that as n increases, the \mathcal{BBS} is chosen more frequently.

Table 12 contains the results when H_g is the true hypothesis. Once again, the null rejection rates stayed close of the nominal levels. When $n = 30$ and $\epsilon = 0.05$, adding the cells relative to $R2$ and $R4$, we

see that the null rejection rate of H_g is 5.38%, a value close of the nominal level considered. Moreover, we could observe that the results of Table 12 were better than when H_f is the true hypothesis. In the column corresponding to $R3$ are presented the proportion of time that only the \mathcal{GBS}_2 is chosen as the most adequate model without considering the sign of W_{ne} . When $\epsilon = 0.05$, for $n = 30$ this happens 28.04% of time and for $n = 150$ it happens 93.44% of time, a superior performance of when H_f is true. This can also be seen in the sixth column, where we can observe that the proportion of time the \mathcal{GBS}_2 distribution is correctly chosen are higher than the corresponding values in the fifth column of Table 11.

Therefore, we can note from these results that the nonnested hypothesis tests with bootstrap proposed for the \mathcal{BS} and \mathcal{GBS}_2 models presented satisfactory performances. For both distributions the null rejection rates are close of the nominal levels when the generation is made under the correct null hypothesis. We can also observe that, as n increases, the tests tend to indicate the true distribution at higher rates. We also noted that, for the models considered, the tests presented better performance when the true model is the \mathcal{GBS}_2 distribution.

7. EMPIRICAL APPLICATIONS

7.1. Runoff amounts. We shall now return for the data set remarked on Section 1, which we used to illustrate how the problem of non convergence of optimization processes during the parameters estimation can occur in practical situations with the \mathcal{BS} model. The data, provided by Folks and Chhikara (1978), consists of 25 runoff amounts at Jug Bridge, in Maryland. Table 13 contains some descriptive statistics of this data set. We can observe that the data has a large kurtosis coefficient, i.e., it has a leptokurtic distribution, and has low variance, which might indicate the data is concentrated around the mean and median values; these characteristics might be an indicative we are dealing with a unimodal data set.

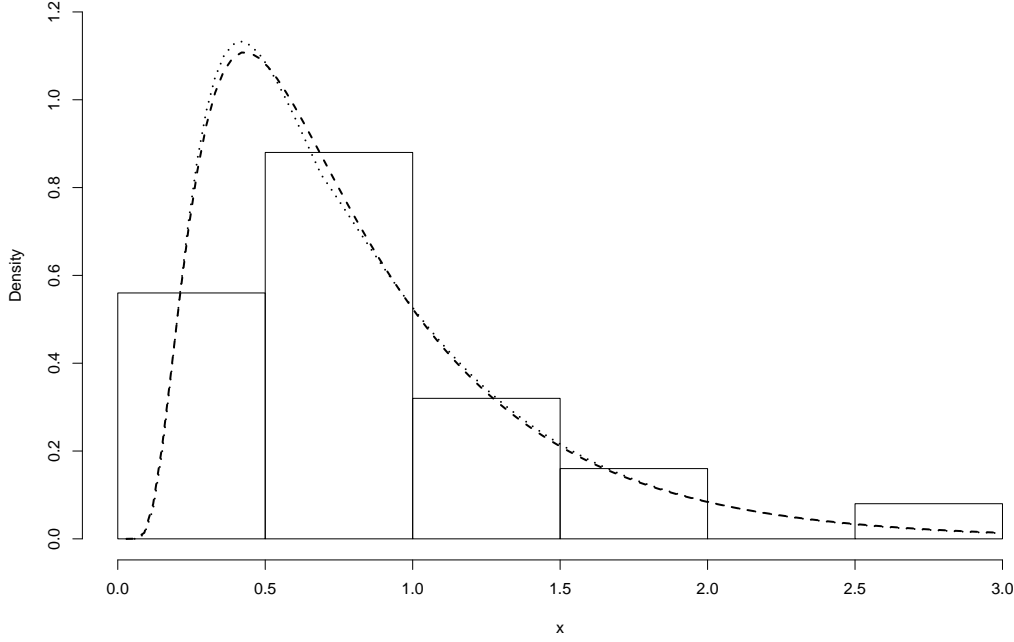
TABLE 13. Descriptive statistics from the runoff data.

| min | max | median | mean | variance | asymmetry | kurtosis |
|------|------|--------|------|----------|-----------|----------|
| 0.17 | 2.92 | 0.7 | 0.84 | 0.3459 | 1.7953 | 6.7493 |

The models fitted were the $\mathcal{BS}(\alpha, \beta)$ and $\mathcal{GBS}(\alpha, \beta, \gamma)$ distributions. The parameter estimates of the first model were $\hat{\alpha} = 0.66$ (0.0936) and $\hat{\beta} = 0.69$ (0.0865), with standard errors in parenthesis. For the second model the maximum likelihood estimates were not possible to be obtained because the optimization process failed to reach convergence. As was showed in Figure 1a, the log-likelihood have a region apparently flat for some values of the parameter α and γ , with the value of β being fixed at 0.69. Nonetheless, using the estimator MLE_p the process reaches convergence and, as we can see in Figure 1b, the penalization modified the log-likelihood function in a way that a maximum point is possible to be reached by the optimization process. The parameters estimates of the \mathcal{GBS} model using the MLE_p were $\hat{\alpha} = 0.63$ (0.2287), $\hat{\beta} = 0.69$ (0.0817) and $\hat{\gamma} = -0.13$ (0.8449). We can observe that the standard error of $\hat{\gamma}$ is large relative to the point estimate and indicates the \mathcal{BS} model is more adequate.

Figure 10 contains the histogram of the data set and the fitted densities. We can note that the estimates obtained for the models lead to very similar densities. Since the \mathcal{BS} distribution is simpler than the \mathcal{GBS} distribution, it seems more suitable than the bimodal distribution. To evaluate this, we tested the hypothesis $H_0 : \gamma = 0$ against a two-sided alternative. The p -values of the tests LR, score,

FIGURE 10. Histogram of the runoff data with the fitted densities obtained with $\mathcal{BS}(0.66, 0.69)$ (dashed line) and $\mathcal{BBS}(0.63, 0.69, -0.13)$ (dotted line).



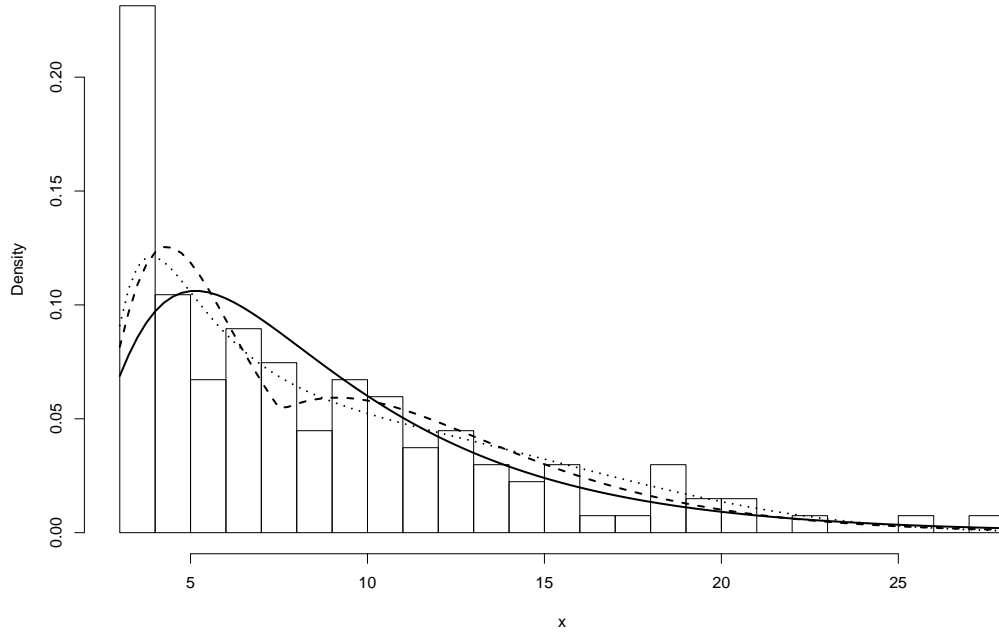
Wald, LR_{pb} , LR_{bbc} and S_{pb} were 0.85, 0.81, 0.87, 0.92, 0.89 and 0.91, respectively. Therefore, we have strong evidences that the fit obtained with the \mathcal{BS} distribution is better.

In summary, this example illustrates how is possible to find cases where the optimization process being used to obtain maximum likelihood estimates in the \mathcal{BBS} model fails to reach convergence, but that might be possible to solve this by using a penalized log-likelihood function, allowing to carry on further analysis that might indicate whether the \mathcal{BBS} is suitable or the \mathcal{BS} provides a better fit.

7.2. Depressive condition data. The second application example is a data set about 134 children emotional condition, with their measures of depression. The data was analyzed, for example, by Leiva et al. (2010) and Balakrishnan et al. (2011), where both works involved mixture of distribution in the analysis.

Table 14 presents some descriptive statistics of this data set, where we can see the data is right-skewed, has leptokurtic distribution and quite large variance. The fitted model were \mathcal{BS} , \mathcal{BBS} and \mathcal{GBS}_2 , obtaining the Akaike (AIC) and Schwarz (BIC) information criteria for each fit. The estimates, with standard error in parenthesis, of the parameters from the \mathcal{BS} model were $\hat{\alpha} = 0.603$ (0.0368) and $\hat{\beta} = 7.58$ (0.3773), providing AIC and BIC values of 780.09 and 785.89, respectively. The parameters estimates from the \mathcal{BBS} model were $\hat{\alpha} = 0.42$ (0.0481), $\hat{\beta} = 7.54$ (0.2645) and $\hat{\gamma} = -0.85$ (0.2569), with AIC and BIC values of 776.26 and 784.95, respectively. For the \mathcal{GBS}_2 model the estimates were $\hat{\alpha} = 2.38$ (0.6290), $\hat{\beta} = 7.74$ (0.3045) and $\hat{\nu} = 1.53$ (0.2525), with AIC and BIC values of 771.78 and 780.47, respectively. The histogram of the data set with the fitted densities of these models are presented in Figure 11.

FIGURE 11. Histogram of the depressive condition data with the fitted densities obtained with $\mathcal{BS}(0.60, 7.58)$ (solid line), $\mathcal{BBS}(0.42, 7.54, -0.85)$ (dashed line) and $\mathcal{GBS}_2(2.38, 7.74, 1.53)$ (dotted line).



From the information criteria, we would conclude that the most adequate model is the \mathcal{GBS}_2 distribution, followed by the \mathcal{BBS} distribution. To test which of these two models is the most suitable for this data, a nonnested hypothesis test was performed, providing a test statistic of value $W_{ne} = -0.0167$ with p -value of 0.0189 under H_f (supposing the \mathcal{BBS} distribution is the true model) and p -value of 0.6453 under H_g (supposing the \mathcal{GBS}_2 is the true model). Then, we conclude that the \mathcal{GBS}_2 is more adequate for the emotional condition data set.

TABLE 14. Descriptive statistics of the depressive condition data.

| min | max | median | mean | variance | asymmetry | kurtosis |
|-----|-----|--------|------|----------|-----------|----------|
| 3 | 28 | 8 | 8.96 | 28.73 | 1.11 | 3.88 |

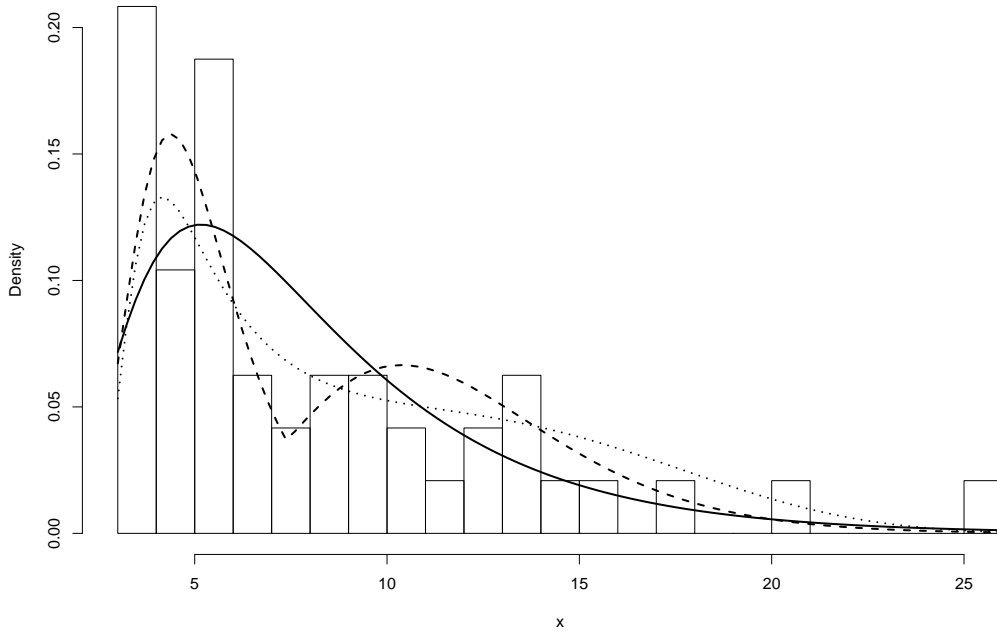
7.3. Adhesive strength. The third data set analyzed is provided by Ehsani et al. (1996) and was also analyzed by Olmos et al. (2016), who used the \mathcal{BBS} distribution in the analysis. The data is consisted of 48 observations about adhesive strength to concrete of bars reinforced with glass fibre. Some descriptive statistics are presented in Table 15, where we can observe that the data presents high kurtosis coefficient, with value greater than 5, is right-skewed and has a variance value much greater than the mean and median.

Once more, the fitted models were \mathcal{BS} , \mathcal{BBS} and \mathcal{GBS}_2 . For the first model the parameters estimates were $\hat{\alpha} = 0.54$ (0.0553) and $\hat{\beta} = 7.05$ (0.5316), providing AIC and BIC values of 264.52 and

TABLE 15. Descriptive statistics of the adhesive strength data.

| min | max | median | mean | variance | asymmetry | kurtosis |
|-----|------|--------|------|----------|-----------|----------|
| 3.4 | 25.5 | 5.95 | 8.08 | 23.7017 | 1.448 | 5.0345 |

FIGURE 12. Histogram of the adhesive strength data with the fitted densities obtained with $\mathcal{BS}(0.54, 7.05)$ (solid line), $\mathcal{BBS}(0.31, 7.39, -1.38)$ (dashed line) and $\mathcal{GBS}_2(3.19, 8.05, 1.99)$ (dotted line).



268.26, respectively. The parameter estimates of the second model were $\hat{\alpha} = 0.31$ (0.0460), $\hat{\beta} = 7.39$ (0.3162) and $\hat{\gamma} = -1.38$ (0.3525), providing AIC and BIC values of 260.06 and 265.67, respectively. For the \mathcal{GBS}_2 model, the parameter estimates were $\hat{\alpha} = 3.19$ (1.5536), $\hat{\beta} = 8.05$ (0.5371) and $\hat{\gamma} = 1.99$ (0.5203), providing AIC and BIC values of 262.26 and 267.88, respectively. The histogram of the fitted densities is shown in Figure 12.

For this data set, the best fit according to the information criteria was the \mathcal{BBS} distribution, followed by the \mathcal{GBS}_2 model. The nonnested hypothesis test of these two models provided a test statistic of value $W_{ne} = 0.0229$, with p -value of 0.6543 under H_f and with p -value of 0.0489 under H_g , and then, we can see that there is significant evidence that the \mathcal{BBS} distribution is the most adequate model for this data.

Adopting the \mathcal{BBS} as the most suitable model, tests to verify if the data has a bimodal distribution were done. The hypothesis tested were $H_0 : \gamma \geq 0$ and $H_1 : \gamma < 0$. The p -values of the tests SLR, SLR_{c1} , SLR_{c2} and SLR_{bp} were 0.0002, 0.0007, 0.0006 and 0.002, respectively, i.e., in all tests we reject H_0 in favor of H_1 . Hence, there is strong evidences that $\gamma < 0$, which indicates the data comes from a bimodal distribution.

8. CONCLUDING REMARKS

Optimization processes might fail to reach convergence with considerable frequency during maximum likelihood estimation in the \mathcal{BBS} model. A penalization in the log-likelihood was proposed using a modification of the Jeffreys prior. Different methods were tried to solve the non convergence problems and the ones that use some penalization in the log-likelihood presented the lowest rates of non convergence. In general, the log-likelihood penalized with the modified Jeffreys prior presented the best performance, considering the trade off between quality of the estimates and the non convergence rates of the methods analyzed.

Hypothesis tests with the \mathcal{BBS} model were studied with the penalized log-likelihood, using the modified Jeffreys prior penalization. We analyzed the likelihood ratio, score and Wald tests, observing that the tests are quite liberal, but that their null rejection rates tend to the nominal levels as the sample size increases. The first two tests had the best performances, and we could note that bootstrap resampling can provide much better results. One-sided tests using the signed likelihood ratio were also investigated. We compared the sizes and powers of the test without corrections, with analytical and with bootstrap corrections; the tests with analytical corrections presented the best performances in terms of size and power. Nonnested tests with bootstrap were also considered, providing a way to choose between the \mathcal{BBS} model and another version of the Birnbaum-Saunders distribution that exhibits bimodality. Since in this case there are two distributions, the test was held considering two null hypothesis at time. Besides this difficulty, we could note that null rejection rates of the true hypothesis gets closer to the true nominal level as the sample size increases and that the test tends to indicate the true model more frequently.

An example with real data where the optimization process fails to reach convergence during estimation with the \mathcal{BBS} model was presented, illustrating how this problem can appear in practice and that plausible estimates can be obtained using a penalized log-likelihood. Another two examples with real data were discussed, with distinct results for the nonnested hypothesis tests. For the data where the \mathcal{BBS} distribution is chosen as the most suitable model, unilateral tests were performed to verify if their distribution is unimodal, proving results compatible with the empirical distribution of the data set.

ACKNOWLEDGEMENTS

We gratefully acknowledge partial financial support from CAPES and CNPq.

REFERENCES

- A. Azzalini and R. B. Arellano-Valle. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143(2):419–433, 2013.
- N. Balakrishnan, V. Leiva, A. Sanhueza, and F. Vilca. Estimation in the Birnbaum–Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Sort*, 33(2):171–192, 2009.
- N. Balakrishnan, R. C. Gupta, D. Kundu, V. Leiva, and A. Sanhueza. On some mixture models based on the Birnbaum–Saunders distribution and associated inference. *Journal of Statistical Planning and Inference*, 141(7):2175–2190, 2011.
- O. E. Barndorff-Nielsen. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, 73(2):307–322, 1986.
- O. E. Barndorff-Nielsen. Modified signed log likelihood ratio. *Biometrika*, 78(3):557–563, 1991.

- Z. W. Birnbaum and S. C. Saunders. A new family of life distributions. *Journal of Applied Probability*, 6:319–327, 1969a.
- Z. W. Birnbaum and S. C. Saunders. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6:328–347, 1969b.
- M. Bourguignon, R. B. Silva, and G. M. Cordeiro. A new class of fatigue life distributions. *Journal of Statistical Computation and Simulation*, 84(12):2619–2635, 2014.
- G. M. Cordeiro and F. Cribari-Neto. *An Introduction to Bartlett Correction and Bias Reduction*. Springer, New York, 2014.
- G. M. Cordeiro and A. J. Lemonte. The exponentiated generalized Birnbaum–Saunders distribution. *Applied Mathematics and Computation*, 247:762–779, 2014.
- G. M. Cordeiro, A. J. Lemonte, and E. M. Ortega. An extended fatigue life distribution. *Statistics*, 47(3): 626–653, 2013.
- D. R. Cox. Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 105–123, London, 1961. Birckbeck College.
- D. R. Cox. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B*, 24:406–424, 1962.
- F. Cribari-Neto, A. C. Frery, and M. F. Silva. Improved estimation of clutter properties in speckled imagery. *Computational Statistics & Data Analysis*, 40(4):801–824, 2002.
- A. H. Cysneiros, F. Cribari-Neto, and C. A. Araújo. On Birnbaum-Saunders inference. *Computational Statistics & Data Analysis*, 52(11):4939–4950, 2008.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, New York, 1997.
- J. A. Díaz-García and J. R. Dominguez-Molina. Some generalisations of Birnbaum-Saunders and sinh-normal distributions. *International Mathematical Forum*, 1(35):1709–1727, 2006.
- J. A. Díaz-García and V. Leiva. A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128(2):445–457, 2005.
- T. J. DiCiccio and M. A. Martin. Simple modifications for signed roots of likelihood ratio statistics. *Journal of the Royal Statistical Society B*, 55:305–316, 1993.
- J. A. Doornik. *An Object-Oriented Matrix Programming Language Ox 6*. Timberlake Consultants Press, London, 2009.
- B. Efron. More efficient bootstrap computations. *Journal of the American Statistical Association*, 85(409):79–89, 1990.
- M. R. Ehsani, H. Saadatmanesh, and S. Tao. Design recommendations for bond of gfrp rebars to concrete. *Journal of Structural Engineering*, 122(3):247–254, 1996.
- S. L. Ferrari and E. C. Pinheiro. Small-sample one-sided testing in extreme value regression models. *Advances in Statistical Analysis*, 100:79–97, 2016.
- D. Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38, 1993.
- J. Folks and R. Chhikara. The inverse gaussian distribution and its statistical application—a review. *Journal of the Royal Statistical Society B*, 40(3):263–289, 1978.
- D. A. S. Fraser, N. Reid, and J. Wu. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, 86(2):249–264, 1999.
- H. W. Gómez, D. Elal-Olivero, H. S. Salinas, and H. Bolfarine. Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics*, 22(1):50–62, 2011.
- V. Leiva. *The Birnbaum-Saunders Distribution*. Academic Press, London, 2015.

- V. Leiva, A. Sanhueza, S. Kotz, and N. Araneda. A unified mixture model based on the inverse gaussian distribution. *Pakistan Journal of Statistics*, 26(3):445–460, 2010.
- V. Leiva, M. Tejo, P. Guiraud, O. Schmachtenberg, P. Orio, and F. Marmolejo-Ramos. Modeling neural activity with cumulative damage distributions. *Biological cybernetics*, 109:421–433, 2015.
- A. J. Lemonte and S. L. Ferrari. Signed likelihood ratio tests in the Birnbaum–Saunders regression model. *Journal of Statistical Planning and Inference*, 141:1031–1040, 2011.
- A. J. Lemonte, F. Cribari-Neto, and K. L. P. Vasconcellos. Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 51(9):4656–4681, 2007.
- A. J. Lemonte, A. B. Simas, and F. Cribari-Neto. Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution. *Journal of Statistical Computation and Simulation*, 78(1):37–49, 2008.
- F. Lewis, A. Butler, and L. Gilbert. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2):155–162, 2011.
- B. Liseo. La classe delle densità normali sghembe: aspetti inferenziali da un punto di vista Bayesiano. *Statistica*, 50(1):71–79, 1990.
- R. C. Mittelhammer, G. G. Judge, and D. J. Miller. *Econometric Foundations*. Cambridge University Press, New York, 2000.
- H. Ng, D. Kundu, and N. Balakrishnan. Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 43(3):283–298, 2003.
- N. Olmos, G. Martinez-Florez, and H. Bolfarine. Bimodal Birnbaum–Saunders distribution with applications to non-negative measurements. *Communications in Statistics–Theory and Methods*, 2016.
- W. J. Owen. A new three-parameter extension to the Birnbaum–Saunders distribution. *IEEE Transactions on Reliability*, 55(3):475–479, 2006.
- W. J. Owen and H. K. T. Ng. Revisit of relationships and models for the Birnbaum–Saunders and inverse-gaussian distributions. *Journal of Statistical Distributions and Applications*, 2(1):1–23, 2015.
- A. G. Patriota. On scale-mixture Birnbaum–Saunders distributions. *Journal of Statistical Planning and Inference*, 142(7):2221–2226, 2012.
- D. M. Pianto and F. Cribari-Neto. Dealing with monotone likelihood in a model for speckled data. *Computational Statistics & Data Analysis*, 55(3):1394–1409, 2011.
- D. M. Rocke. Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*, 84(406):598–601, 1989.
- A. Sanhueza, V. Leiva, and N. Balakrishnan. The generalized Birnbaum–Saunders distribution and its theory, methodology, and application. *Communications in Statistics–Theory and Methods*, 37(5):645–670, 2008.
- N. Sartori. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 136(12):4259–4275, 2006.
- T. A. Severini. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, 86(2):235–247, 1999.
- T. A. Severini. *Likelihood Methods in Statistics*. Oxford University Press, New York, 2000.
- B. Smith, S. Wang, A. Wong, and X. Zhou. A penalized likelihood approach to parameter estimation with integral reliability constraints. *Entropy*, 17(6):4040–4063, 2015.
- Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333, 1989.

- D. A. Williams. Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. *Biometrics*, 26(1):23–32, 1970.
- J. Wu and A. C. Wong. Improved interval estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 47(4):809–821, 2004.
- X. Zhu and N. Balakrishnan. Birnbaum–Saunders distribution based on Laplace kernel and some properties and inferential issues. *Statistics & Probability Letters*, 101:1–10, 2015.

DEPARTAMENTO DE ESTATÍSTICA, UNIVERSIDADE FEDERAL DE PERNAMBUCO, RECIFE, PE – BRAZIL